Abstract

This paper presents results from five experiments investigating the online interpretation of PRO and pronouns as subjects of adjunct clauses. The experiments attempted to provide evidence on the timecourse of interpretation for PRO/pronouns, and whether this interpretation causes reactivation of the semantic content of the antecedent in memory, using the N400 component of ERPs and self-paced reading times. The results were inconclusive. Although facilitation was seen in the N400 for some conditions, the effect of the interpretation of PRO and pronouns was small and insignificant, even though PRO and the pronoun did affect offline cloze results. Self-paced reading times also did not indicate differences in processing due to PRO/pronoun resolution. These results suggest either that interpreting PRO and pronouns does not reactivate the semantic content of their antecedents sufficient to cause facilitation of related material, or that any such facilitation is not reflected in the N400 or self-paced reading times.

Introduction

What does it mean for an anaphoric dependency to be resolved, and what all occurs in the parser’s memory such that the anaphor is interpreted successfully? This paper investigates the online interpretation of two different kinds of anaphora: the null subject of non-finite adjuncts, which we will refer to as PRO, and unbound pronouns, as illustrated in (1a-b) and (1c), respectively.¹

(1) The copᵢ ignored Jackieⱼ…
   a. . . in order PROᵢ to arrest the thief.
   b. . . while PROᵢ arresting the thief.
   c. . . while heⱼ/sheⱼ arrested the thief.

¹Although we use the PRO label for clarity, our results and conclusions are not dependent on a PRO analysis of the implicit subject of control structures, or even on the presence of a syntactically-represented implicit subject. For a discussion on the debate over the representation of PRO, see Hornstein (2003).
Work based on the Lewis and Vasishth (2005)-based model of memory and processing suggests that a pronoun may have features that correspond to different chunks in memory. These chunks are searched simultaneously for matching features (see also Martin and McElree (2009)). Parker and Phillips (2017) argued that upon encountering a reflexive, people will search for an antecedent based on both the reflexive’s featural and structural requirements simultaneously (i.e. they will search for an argument that both c-commands the reflexive and one that matches in gender), although the structural features may be more important for initial search (Dillon, 2014). For subject-verb agreement, people will search for an NP that matches in number and is also a subject (Wagers, Lau, & Phillips, 2009).

Non-reflexive pronouns, on the other hand, may involve different features in their search for an antecedent. Pronouns have featural requirements for their antecedents, but they generally do not have strong structural requirements. Two exceptions are Condition B, which is a restriction against certain antecedents (Chomsky, 1981), and bound variable readings of pronouns, which often have a c-command requirement (Hornstein & Weinberg, 1990). There are also structural and information-structural biases. A pronoun is more likely to refer to the discourse center, which is often represented in the subject position or some other prominent position of a previous sentence or clause (Grosz, Weinstein, & Joshi, 1995). Indeed, previous literature has provided evidence that structural information does play a role in the on-line resolution of unbound pronouns. In visual-world eye-tracking experiments by Arnold, Eisenband, Brown-Schmidt, and Trueswell (2000), for example, participants were faster to resolve an ambiguous pronoun in sentences such as (2) to the subject of a preceding clause than to the object. Arnold et al. proposed that this is because subjects are more “accessible” antecedents than objects. However, because this bias is not a requirement in the same way a reflexive’s structural features are, it seems likely that structural cues would play a different role in the search for an antecedent to a pronoun.

2 Donald is bringing some mail to Mickey while a violent storm is beginning. He’s carrying an umbrella, and it looks like they’re both going to need it.

Another kind of referential dependency is Control, which underlies the interpretation of the null subject of non-finite clauses (Chomsky, 1981). When the non-finite clause is an adjunct to the matrix clause, this referential dependency is known as Adjunct Control. In this paper I follow the generative tradition of labeling the null subject as PRO, although the discussion here is equally compatible with other theories. What is important is that there is a referential dependency between the understood subject of the embedded clause and some other sentential argument.\footnote{At least in cases of what is known as Obligatory Control. There are also cases of Non-Obligatory Control, where the referential dependency is between PRO and a discourse argument. See Williams (1980) and Landau (2013) for discussion.} Many Control constructions only allow PRO to be Controlled by a
single structurally-determined argument. For cases of Adjunct Control, this is often the subject of the next higher clause (Clark, 1990; Hornstein, 1999; Boeckx, Hornstein, & Nunes, 2010; Parker, Lago, & Phillips, 2015), as illustrated in (3), although this is dependent on the type of adjunct (Landau, 2013).

(3) Jon\textsubscript{i} talked to Mary\textsubscript{j} [before PRO\textsubscript{i}/\textsubscript{j} leaving].

Because the referent of PRO in sentences such as (3) is arguably determined structurally, and because PRO itself lacks any morphological features, the search process for an antecedent to PRO would naturally be different from the search for antecedents to overt pronouns, as it would involve only structural features. Several studies have investigated the online processing of various kinds of Control (e.g. Frazier, Clifton, & Randall, 1983; Nicol & Swinney, 1989; Mauner, Tanenhaus, & Carlson, 1995), including Adjunct Control (Betancort, Carreiras, & Acuña-Fariña, 2006; Parker et al., 2015; McCourt, Green, Lau, & Williams, 2015). These studies have generally found that resolving PRO to the subject of the next higher clause takes longer than resolving it to the object, even when a given construction only allows one interpretation. In addition, resolution of PRO may take longer than resolution of a pronoun (Nicol & Swinney, 1989), although a direct comparison has only been done using cross-modal priming, and Arnold et al. (2000) argued that this method does not accurately reflect the timecourse of interpretation for pronouns. Therefore, it is unclear whether the timecourse of interpretation differs for pronouns and PRO (see Green, McCourt, Lau, & Williams, in preparation, for an eye-tracking study comparing the two).

Although the timecourse evidence from cross-modal priming has been challenged, these studies do provide evidence that pronouns activate the semantic content of their antecedents in memory. Nicol and Swinney (1989) found that both PRO and pronouns can facilitate decision times to words related to their antecedents, although facilitation appeared later for PRO conditions than for pronoun conditions.

The main questions asked in this paper are the following. Does resolution of anaphora involving PRO (as in (4a)) and pronouns (as in (4b)) reactivate the semantic features of the antecedent in memory? If so, will that lead to facilitation of related material following the anaphor? Do these processes differ when the antecedent is the matrix subject versus when it is the matrix object?

(4) The cop\textsubscript{i} ignored Jackie\textsubscript{j} . . .
   a. . . . in order PRO\textsubscript{i} to arrest the thief.
   b. . . . while he\textsubscript{i}/she\textsubscript{j} arrested the thief.

Experiments 1 and 2 investigate these questions for PRO, and Experiments 3, 4, and 5 do so for pronouns.

\(^3\)See also Green (submitted) for a discussion of less-common cases of temporal adjuncts allowing control by a non-subject.
One kind of Adjunct Control involves non-finite rationale clauses, as in (4a). Much of the previous research on Control of rationale clauses has maintained that it involves a syntactic dependency between PRO and its understood antecedent (Roeppe, 1987; Mauner et al., 1995; Whelpton, 1995). In this respect, structural features would be the most important features for a parser to consider when establishing a referential dependency. Others disagree, claiming that discourse factors determine the referent of PRO in rationale clauses (Williams, 1987; Landau, 2000; A. Williams, 2015; A. Williams & Green, 2017). One argument that has been made involves the interpretive similarities between “local” (5a) as compared to “remote” (5b) rationale clauses.

(5)  
   a. The cop ignored Jackie [in order PRO to arrest the thief].  
   b. The cop ignored Jackie. The reason was [PRO to arrest the thief].

Remote rationale clauses differ from local ones in that they are found in a separate sentence from the fact they give the rationale for. Because of this, PRO is in not in the same syntactic domain as its understood Controller, and, arguably, there can be no structural/syntactic dependency between the two (A. Williams, 2015; McCourt et al., 2015; A. Williams & Green, 2017). Finding an antecedent for PRO in remote rationale clauses therefore cannot rely on structural feature search. Using self-paced reading times, McCourt et al. (2015) provide evidence that there is no difference in the processing of local and remote rationale clauses. This result is compatible with at least three interpretations.

First, it is possible that the lack of difference seen between reading times of local and remote rationale clauses was due to their being processed via the same interpretive strategy. In other words, given that remote rationale clauses require some pragmatic/discourse strategy for the interpretation of PRO, a lack of difference in reading times between local and remote conditions could be taken as evidence that local Control of rational clauses is also resolved via pragmatics/discourse. This interpretation, favored by McCourt et al., would counter claims made by Mauner et al. (1995), but would be consistent with the arguments in A. Williams and Green (2017).

The second possible interpretation of the results in McCourt et al. (2015) is that the self-paced reading measure used may not be sensitive enough to detect real differences in the strategy used for processing PRO in local and remote rationale clauses. McCourt et al.’s results are consistent with local Control being a syntactic phenomenon and remote control being a discourse phenomenon if resolving the dependency between PRO and its antecedent in either case does not result in a difference in reading times.

Finally, it could have been the case that PRO was not being interpreted at all in the course of McCourt
et al.’s experiments, or that it wasn’t interpreted soon enough for reading measures to detect. In order to test this possibility, we used a more sensitive measure, EEG, to investigate the processing of rationale clauses. The N400 component in EEG is known to be sensitive to conceptual relations (Lau, Phillips, & Poeppel, 2008), such that words that are conceptually related to previously seen words will have a reduced N400 amplitude. We manipulated whether the verb in local rationale clauses was semantically related to the understood antecedent of PRO, as seen in Table 1.

If PRO is being interpreted as the cop immediately in the PRO conditions, responses at arrest in the PRO-related condition are predicted to result in facilitation compared to responses at catch in the PRO-unrelated condition. This facilitation should exceed any that may result from general priming or prediction due the presence of the cop in the sentence only if PRO reactivates the cop and its semantic content. In other words, the facilitation of arrest in the PRO-related condition, where PRO has reactivated cop just prior, should exceed any facilitation of arrest in the control condition, where the subject of the embedded clause (Bill) is not predicted to reactivate the cop in memory.

### Materials and methods

#### Participants

Participants were recruited at the University of Maryland, and prior written consent was obtained. All participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) and were native speakers of English. There were twenty-nine participants. Five were excluded for excessive artifact (>40% rejection), and 3 more were excluded for low accuracy on the comprehension questions (<80%) for a total of 20 participants in the analysis. Participants were compensated either with course credit or $10/hour (usually 1.5-2 hours total, including prep time).

#### Materials

Sentences were created by combining a finite matrix clause with a rationale clause beginning with in order. Stimuli varied in whether the adjunct was finite and contained an overt subject (the control condition), or

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO-related</td>
<td>The cop left the building in order (PRO) to arrest the thief.</td>
</tr>
<tr>
<td>PRO-unrelated</td>
<td>The cop left the building in order (PRO) to catch the thief.</td>
</tr>
<tr>
<td>control</td>
<td>The cop left the building in order for Bill to arrest the thief.</td>
</tr>
</tbody>
</table>

Table 1: Experiment 1 items. Region of interest underlined.
was non-finite with no overt subject (PRO conditions). Within the PRO conditions, the related/unrelated conditions varied with respect to the relatedness of the verb in the rationale clause to the subject of the matrix clause. The verbs in the related condition were selected by matching nouns with verbs that would be typically associated with that noun. Although this was done based on author intuitions, the association between the noun and verb in each case was verified using Latent Semantic Analysis (Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990).

Ninety sets of items were distributed across three lists in a Latin square design. 180 fillers were also included in each list, for a total of 270 items. One third of the items were followed by comprehension questions. Unrelated verbs in each set were selected from the list of related verbs used in other sets, such that each verb used in a related list was used as the unrelated verb in another set.

Among the noun-verb pairs selected for high semantic association, five verbs were repeated: fight, interview, shoot, steal, and kill. Each participant therefore saw each of these verbs twice in the course of the experiment, which could have led to some repetition facilitation. However, no participant saw the same verb in the same condition, and repeated verbs were never presented in the same or adjacent blocks in an attempt to minimize any such effect.

**Procedure**

Participants sat in a chair while stimuli were presented on a computer screen. Each trial began with a fixation cross that remained in the center of the screen for 200ms, followed by a 200ms blank screen. The sentence was then presented sequentially, word by word, with each word appearing in the center of the screen for 400ms followed by a 200ms blank screen, creating a 600ms inter-stimulus interval. One third of the sentences were followed by a yes-no comprehension question to ensure that participants were paying attention. No feedback was given on their answers.

The experiment began with a practice block of 6 items. Participants then completed a total of 270 items, split into 6 blocks of 45 items each. Participants were encouraged to rest between blocks. The experiment usually lasted 75-90 minutes.

**Electrophysiological recording**

Twenty-nine tin electrodes were held in place on the scalp with an elastic cap (Electro-Cap International, Inc., Eaton, OH) (O1, O2, P7, P3, Pz, P4, P8, TP7, CP3, CPz, CP4, TP8, T7, C3, Cz, C4, T8, FT7, FC3, FCz, FC4, FT8, F7, F3, Fz, F4, F8, FP1, FP2). Bipolar electrodes were placed at the outer canthus of each eye as well as above and below the left eye to monitor horizontal and vertical eye movements. Electrodes were also placed over the left and right mastoids, and scalp electrodes were referenced to them. Impedences were
maintained at less than 10kΩ for scalp electrodes, and at or below 5kΩ for ocular and mastoid electrodes. The EEG signal was amplified using a NeuroScan SynAmps® Model 5083 (NeuroScan, Inc., Charlotte, NC) with a bandpass of 0.05-100 Hz and was continuously sampled at 500 Hz with an analog-to-digital converter.

**Data analysis**

For trials free of artifact, averaged ERPs were formed offline, time-locked to the onset of the embedded verb, using processing routines from the EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes. Participants were excluded if they had artifacts in more than 40% of trials or if their comprehension question accuracy was less than 80%. Within participants that were included, 17% of the total data was excluded for artifacts, and mean accuracy on the comprehension questions was 86.5%. A 40Hz low-pass filter was applied to ERPs, and waveforms were plotted either with a 100ms prestimulus baseline or with the first 100ms post stimulus onset as a baseline.

Because N400 effects often peak at the midline, statistical analysis was performed on the mean ERP amplitude of the five midline electrodes (Fz, FCz, Cz, CPz, Pz) in the 300-500ms after the onset of the embedded verb. Significance was measured using linear mixed effects models in R (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2014), and p-values were obtained using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2014), which uses Satterwaithe approximations to calculate degrees of freedom. Two different models were fitted. The first measured the effect of verb relatedness, comparing the PRO-related and control conditions to the PRO-unrelated condition. The second compared the PRO-related condition to the control condition to determine whether PRO caused any additional facilitation of the related verb. Models included random intercepts for participants, but no random slopes in order to reduce computational cost.

**Results**

Evoked responses are illustrated in Figure 1. Mixed effects models indicated a main effect of verb relatedness on the amplitude of the N400 ($t = -2.472, p < 0.05$) in the midline electrodes; a reduced N400 was seen at the verb in the PRO-related and control conditions compared to the unrelated condition. This effect is illustrated for the PRO condition in Figure 2(a). There was a similar reduction of the N400 at the embedded verb when the subject of the rationale clause was unrelated to the verb. There was no significant difference between the N400 amplitude in the PRO-related condition and the control condition ($t = 1.104, p > 0.1$). In fact, the amplitude of the N400 in the PRO-related condition was numerically higher than in the control condition in midline electrodes, especially in posterior regions (see Figure 2(b)).
Figure 1: Experiment 1: ERPs from scalp electrodes at the non-finite verb of interest. A reduced N400 was seen for both the PRO-related (red) and control (blue) conditions compared to the PRO-unrelated condition (black). There was no significant difference between the PRO-related and control conditions.

Figure 2: Topographical distributions for Experiment 1: Scalp maps demonstrating N400 effects during the 300-500 ms time window. Scalp maps illustrate the mean difference between the PRO-unrelated and PRO-related conditions (a) and between the control and PRO-related conditions (b).
In addition to a reduced N400, the control condition showed increased positivity beginning immediately after the onset of the critical word. Such early differences may be indicative of baseline issues. Because of this, we performed additional analyses after baselining the data from 0 to 100 ms post stimulus onset. A comparison of the original and rebaselined evoked responses for two key electrodes are given in Figure 3, and rebaselined topographical maps are given in Figure 4. After rebaselining, the overall effect of verb relatedness was insignificant ($t = -1.553, p > 0.1$). The difference between the control condition and the PRO-unrelated condition also became insignificant, while the difference between the N400 amplitude between the PRO-related and PRO-unrelated conditions (averaged between 300-500 ms) in the midline electrodes remained marginally significant ($t = -1.871, p = 0.077$). However, although the N400 of the control condition was numerically higher than the PRO-related condition, often overlapping with the PRO-unrelated condition in anterior electrodes, the difference between the control and PRO-related conditions was not significant ($t = -0.980, p > 0.1$).

Discussion

When the data were baselined in the window 100 ms prior to stimulus onset, we found a reduced N400 for both the PRO-related and control conditions in comparison to the PRO-unrelated condition. This suggests that processing a verb that is semantically related to the matrix subject was easier than processing an unrelated verb. However, we found no significant difference between the PRO-related and control conditions, which suggests that the reduction in the N400 amplitude for related verbs compared to unrelated verbs was not due
(soley) to reactivation of the matrix subject at PRO. When participants read the unrelated name (e.g. Bill), they presumably were not reactivating the matrix subject (cop). They may have reactivated the matrix subject at PRO, but since there was no difference between the PRO and control conditions, it does not appear to be the case that reactivation of the matrix subject at PRO is what is driving the N400 reduction for the PRO-related and PRO-unrelated conditions.

**Baselining concerns**

When baselined in the 100 ms window prior to the onset of the critical word, the results indicate no difference between the PRO-related and control conditions in N400 amplitude. However, the control condition was more positive than both the PRO-related and PRO-unrelated conditions immediately after the critical word onset (see Figure 1). This may be indicative of baseline distortion. Baselining is successful when the stimuli are identical up to the critical word. However, in this experiment, although the word prior to the critical word was identical in all three conditions (‘to’), the control condition differed prior to that point from the other two conditions by the addition of two words: ‘for’, and a name representing a new discourse referent. Introducing new discourse referents can cause an increased amplitude in the N400 (Burkhardt & Roehm, 2007). Because of this, it is possible that there was a sustained negativity in the control condition up to the critical word, and that baselining lowered the response in the critical window immediately after the critical word onset, inducing an artificial reduction of the N400.

When the results are instead baselined at 0–100 ms post stimulus onset, thus reducing early, potentially...
artifactual differences between the conditions, the N400 of the control condition trends higher than the PRO-related condition, especially in anterior electrodes (see Figures 3 and 4). Although this trend does not result in a significant difference between the PRO-related and the control conditions, it is suggestive of an effect of PRO: specifically that the antecedent of PRO is activated in memory by the critical word such that the N400 of a related verb is reduced.

Our justification for re-baselining at a later window is that the introduction of a new referent in the control condition, but not the other conditions, may have caused the overall amplitude of the response to be more negative. Upon baselining, the response in this condition would be shifted in the positive direction, which may have artifactually lowered the N400 response. However, it is difficult to determine what effect the addition of the new discourse referent had, since there is no comparable region in the other conditions. The best that we can do is to look at the response at ‘to’, the word immediately preceding the critical word. This comparison is given in Figure 5. The response at ‘to’ shows a decreased N400 amplitude in the control condition only that is sustained through the critical word. This may lend support to the idea that baselining artificially lowered the response in the control condition. We tentatively conclude from this that baselining at a window closer to the N400 response of interest (0–100 ms after the onset of presentation of the critical-word) is justified.
Limited affect of PRO

Even if baselining in the 0–100 ms window post stimulus onset is justified, the effect of the PRO-related condition in comparison to the control condition is small, and limited to central, anterior electrodes. In posterior electrodes, the N400 was still just as reduced in the control condition as in the PRO-related condition. One possible explanation for this is that the matrix subject (the cop in the items given in Table 1) may have been so salient in the sentence that the related verb (e.g. arrest) was either predicted or primed in both PRO and control conditions. This would explain why the control condition also showed a reduced N400 compared to the unrelated condition. This alone, though, is not enough to explain a lack of difference between the PRO and control conditions. It is unlikely that we are seeing a floor effect due to this priming, since these N400s are still relatively high. It may also have been the case, though, that PRO was not interpreted immediately (i.e. the to in the embedded clause was not enough to cause a Control dependency to be established by the non-finite verb). Combined with a general priming explanation of the small reduction in the N400 seen, this would explain the results seen, and it would be consistent with earlier research on the processing of Control that suggests it is slower than the interpretation of an overt pronoun (Nicol & Swinney, 1989).

Second, in Experiment 1, PRO and its understood antecedent were always in subject position. Because subjects are more likely to be talked about again in later discourse (Grosz et al., 1995), it is possible that participants were keeping the matrix subject active in memory to some extent. This could explain why the related verb in the rationale clause showed a reduced N400 regardless of whether the rationale clause’s subject was PRO or Bill. It does not seem reasonable that any and all subjects are maintained active in memory throughout the entire discourse, but the stimulus items in this experiment consisted solely of single-sentence items. It would be sufficient for this type of explanation, then, if subjects (or any referent, regardless of subjecthood) remain active during the processing of a single sentence.

Another possible reason the related subject may have been active in memory in both PRO and control conditions is that this experiment used in order clauses, which involve a strong semantic dependency between the two clauses. Specifically, in order in a rationale clause indicates that the party responsible for the fact to be explained is important (A. Williams, 2015; A. Williams & Green, 2017), which is usually the subject of the preceding clause. Therefore, in order in our items may have indicated to participants that the matrix clause subject would be relevant for the interpretation of the upcoming clause. The matrix subject may thus be reactivated at in order, which would lead to facilitation in either the PRO-related or control conditions.

These possible explanations of the results from Experiment 1 leave several questions unanswered: (i) Were participants predicting the related verb in all conditions, regardless of whether the subject of the rationale
Table 2: Sample items for Experiment 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO</td>
<td>The cop left the building in order to __________________________.</td>
</tr>
<tr>
<td>control</td>
<td>The cop left the building in order for Bill to ______________________.</td>
</tr>
</tbody>
</table>

Experiment 2

One potential explanation for the reduced N400 seen in both the PRO-related and the control conditions in Experiment 1 is that the related verb may have been equally predicted in both cases, despite the differences in the embedded clause. This possibility was tested in a cloze task on Mechanical Turk. The items for the cloze task were created from Experiment 1 items, with everything after to removed. A sample pair of items is given in Table 2. Participants (60 native speakers of American English) were asked to finish the sentence as quickly as possible. The task took roughly 45 minutes, and participants were compensated $4.50.

Results

The results of this experiment are given in Figure 6. A chi squared contingency test indicated that the related verb was significantly more predicted in the PRO items than in the control items (16.9% vs. 4.8%, $\chi^2 = 208.3, p < 0.001$).
Discussion

The cloze probability of the related verb used in Experiment 1 was significantly higher in the PRO condition than the control condition in Experiment 2. Therefore, the lack of a difference in N400 amplitude for the PRO-related and control conditions in Experiment 1 was not likely due to arrest being equally predicted in the two conditions, assuming that the tasks in the two experiments did not lead participants to make different predictions. However, although there is a significant difference between the cloze probability of the verb in the two conditions, these probabilities are still both quite small. Therefore, even if a difference in predictability were evident in the N400, the effect may still be a small one.

It is still possible either that arrest was equally primed in the two conditions, if not equally predicted, or that PRO was not being interpreted quickly. If arrest was simply being primed in all of the conditions in Experiment 1, this may have been due to at least two aspects of the design. First, part of the meaning of rationale clauses involves the notion of the party responsible for the fact to be explained (A. Williams, 2015; A. Williams & Green, 2017). The “responsible party” in Experiment 1’s items was the matrix subject, even in the control condition. Because of this, when participants reached in order in any of the conditions, they may have activated the matrix subject as the responsible party. Thus, even if the subject of the rationale clause was not PRO, the matrix subject (e.g. the cop) may have been activated recently, causing facilitation of the related verb (e.g. arrest). Another possibility is that because subjects are more likely to be referred to again in coming discourse, the referent of the matrix subject is somehow maintained in memory, such that words related to it will always be primed. Finally, it is also possible that not only the subject of the matrix clause, but the rest of its content led to priming of the related verb. For example, one set of items was (6). Here, not only is audition semantically related to actor, but it is also related to studio. It may have been
the case that the entire content of the matrix clause primed \textit{arrest} enough to reduce the N400 amplitude in both the PRO-related and control conditions (see Ettinger, Feldman, Resnik, and Phillips (2016)).

(6) The actor went into the studio in order
   a. to audition/beg for a part in the show.
   b. for John to audition for a part in his show.

Under any of these possible scenarios (or a combination of them), general priming of \textit{arrest} would have led to a reduced N400 in the control conditions. It is possible that PRO was still interpreted quickly and caused activation of its antecedent, but that this activation did not cause facilitation of the related verb beyond that caused by general priming.

**Experiment 3**

The results from Experiment 1 showed only a small (and statistically insignificant) difference in the N400 response to a non-finite verb related to the matrix subject when the preceding word was coreferent with the matrix subject (PRO conditions) compared to when it was not (control condition). In addition, this small difference was only present in a few anterior electrodes when the responses were baselined in the 100 ms post stimulus onset. In all other electrodes, there was no difference between the two conditions. As discussed above, two possible explanations for this result are that the matrix clause may have contained material that primed the related verb generally (see (6)), or that \textit{in order} caused (re)activation of the matrix subject due to the meaning of rationale clauses. Experiment 3 removed those two potential confounding factors. In addition, this experiment used overt pronouns instead of PRO in order to ensure that a differential priming effect would be seen with a more obvious referential expression, the idea being that if even overt pronouns do not cause a differential effect, then it is unlikely that null PRO would do so.

This experiment used the N400 to investigate the extent to which a pronoun in a temporal adjunct clause reactivates semantic features of its antecedent. Participants were presented with sentences containing a temporal adjunct clause, as illustrated in Table 3. The subject of the adjunct clause was either a pronoun that was intended to be understood as coreferential with the subject of the matrix clause, or an unrelated name. The verb in the adjunct clause was semantically related to the matrix subject. If the pronoun activates semantic features of its antecedent, we would expect to see facilitation in the N400 component at the verb underlined in Table 3 in the pronoun condition. This facilitation should be greater than in the control condition, where the name in the adjunct clause was not predicted to activate features of the matrix subject.
Table 3: Experiment 3 sample experimental items. Region of interest underlined.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>The actor thought about taxes while he lazily auditioned for a part in the show.</td>
</tr>
<tr>
<td>control</td>
<td>The actor thought about taxes while Emily lazily auditioned for a part in the show.</td>
</tr>
</tbody>
</table>

Materials and methods

Participants

Sixteen participants were recruited at the University of Maryland, and written consent was obtained. All participants were right-handed native speakers of English. No participants were excluded for low accuracy on comprehension questions (all were >80%) or for high artifact (all participants had <40% rejection rates). Participants were compensated either with course credit or with $10/hour (usually 1.5 hours total).

Materials

Stimuli consisted of 60 sets of sentences of two conditions each. All sentences contained a matrix and a temporal adjunct clause headed by before (14 sets), after (20 sets), or while (26 sets), and followed by either a pronoun or a name, an adverb, and a verb semantically related to the matrix subject. Verbs were selected for relatedness to target nouns based on the authors’ intuitions and confirmed by computing the cosine similarity of the two words using Latent Semantic Analysis, as in Experiment 1. The content of the first clause was designed to be specifically non-predictive of the verb in the adjunct clause. This was done to reduce the chance of facilitation of the verb simply due to general priming or predictability, as may have been the case in Experiment 1. Stimuli were distributed across two lists using a latin square design. Each list also had 120 filler items for a total of 180 items per list. The order of experimental items and fillers in each list was randomized.

Procedure

The procedure for Experiment 3 largely followed that of Experiment 1, including impedance thresholds. One participant, though, had a higher impedance of 12.9 for one scalp electrode only (O2). The rest of the procedure differed from Experiment 1 only in the following ways. The 180 items were divided into

---

The adverb was included because Experiment 3 was a pilot intended to evolve into an Adjunct Control experiment involving sentences such as The actor thought about taxes while lazily auditioning for a part in the show. In this sentence, the first cue that the sentence involves a Control structure is the adverb. Without it, the first cue would be the verb itself, and it is unclear/unlikely that this would have given enough time to activate the features of the antecedent to PRO sufficiently for facilitation of the verb itself. The adverb was intended to give that cue one word before the verb. Since we failed to see any extra facilitation in the pronoun condition in this experiment, the Adjunct Control version was not completed. But it may have been the case that the adverb put the pronoun and verb too far away for such facilitation to take place.
shorter blocks, with 6 blocks of only 30 items each, with a break between each block. The format of the comprehension questions that followed one-third of the items also differed from Experiment 1. Instead of yes/no questions, participants answered 2-way multiple choice questions, and they received feedback at the end of each block about their accuracy during that block. In addition, the inter-stimulus interval was reduced to 500 ms. These differences were intended to encourage greater comprehension and participant alertness. Presentation of the materials lasted 45-60 minutes.

**Electrophysiological recording and data analysis**

Recording procedures were the same as in Experiment 1. No subjects were excluded due to high artifact or low comprehension question accuracy. A total of 13.2% of the data was excluded due to artifact. Mean question accuracy was 95.1%. Statistical analysis was performed using linear mixed effects models on the mean ERP amplitude from the five midline electrodes. A single model compared the two conditions with participants as random effects.

**Results**

ERPs for Experiment 3 are given in Figure 7, with a topographic map in Figure 8. No significant difference \((t = 0.642, p > 0.1)\) was seen between the two conditions at the embedded verb (the underlined item in Table 3). However, as in Experiment 1, there was an increased negativity in the control condition beginning immediately after the onset of the critical word, again suggestive of potential baseline distortion. A comparison of ERPs from two key electrodes with the original (prestimulus) and modified (0–100 ms post stimulus onset) baselines is given in Figure 9, and a rebaselined topographic map is given in Figure 10. After re-baselining, there is a small numerical difference between the control and pronoun conditions, with a slightly larger amplitude in the N400 for the control condition, but this difference remains statistically non-significant \((t = -0.884, p > 0.1)\).

**Discussion**

In Experiment 3, it was predicted that if the pronoun activated the semantic features of its antecedent, it would cause facilitation of the verb. Thus the N400 in the pronoun condition was predicted to be smaller than in the control condition. As can be seen in Figures 7 and 8, when the ERPs are baselined in the pre-stimulus window, there is no difference in the N400 in the two conditions. However, as in Experiment 1, it is possible that the introduction of a new discourse referent two words prior to the critical word in the
Figure 7: Experiment 3: ERPs at the verb of interest. There was no difference in the N400 for the pronoun (black) and control (red) conditions.

Figure 8: Topographical distribution for Experiment 3: Scalp maps demonstrating N400 effects during the 300-500 ms time window. Scalp maps illustrate the mean difference between the control and pronoun conditions.
Figure 9: Baseline effects in Experiment 3. ERPs from FCZ and CPZ are given for the original baseline procedure (baselined 0–100 ms pre-stimulus onset) and for the modified baseline (0–100 post stimulus onset).

Figure 10: Topographical distribution for Experiment 3: Scalp maps demonstrating N400 effects during the 300-500 ms time window, baselined at 0–100 ms post stimulus onset.
control condition, but not the pronoun condition, may have led to distortion of the baseline, as is suggested by the increased positivity in the control condition immediately after the onset of the critical word.

**Re-baselining**

If the ERPs are re-baselined in the 100 ms window after the onset of the stimulus, as illustrated in Figures 9 and 10, we find a slight numerical decrease in the N400 in the pronoun condition compared to the control condition. Although this effect was statistically non-significant, it would suggest that the pronoun is being successfully interpreted and that this interpretation facilitated the processing of the upcoming word to a small extent.

Unlike in Experiment 1, where it was not possible to directly examine the effect of introducing a discourse referent, in this experiment, the name in the control condition was matched in the pronoun condition. If the introduction of a new discourse referent were causing a distortion of the baseline in the critical region, then we would expect there to be a sustained difference between the two conditions starting at the name/pronoun region and extending to the critical region. As can be seen in Figure 11, introduction of a new discourse referent by means of a proper name did cause an increased N400, as well as a later positivity (P600), consistent with previous findings (Burkhardt & Roehm, 2007; Burkhardt, 2007). However, these effects were not sustained. By the end of the intervening adverb (i.e. before the start of the critical word), the two conditions are again matched in most electrodes, including many of those where the largest difference in the two conditions appeared at the critical word after re-baselining (e.g. C3 and T7). It is therefore unclear whether re-baselining after the critical word onset is justified.

**Possible explanations for lack of difference**

Since there appears to be no strong justification for baselining after the critical word onset, we must conclude that there was no difference between the pronoun and control conditions. And as in Experiment 1, even in the re-baselined data, the effect of the pronoun was small. Because this experiment did not include a control condition with an unrelated verb, this null effect either reflects equal facilitation of the verb in both conditions, or a lack of facilitation in either condition. I will discuss both possibilities below.

First, the lack of difference in N400 amplitude may have been due to equal facilitation occurring in both conditions. There are several potential reasons that this might be the case. Although the cloze task in Experiment 4 shows that the related verb was not equally predicted in the two conditions, the stimuli were designed such that the matrix clause would be as unrelated as possible to the matrix subject with the intent

---

5 This was again due to its being a pilot experiment.
that the related verb would not be predicted. Therefore, the cloze probability of the related verb was low enough that it may have been unlikely to result in a reduced N400 amplitude.

Another potential cause of equal facilitation in the two conditions might be persistent activation of the matrix subject. Again, because subjects are much more likely to be referred to again in later discourse, it is possible that comprehenders keep the subject active in memory throughout the sentence. For the items in Table 3, for example, *the actor* would remain active in memory and facilitate the related word *audition* within the same sentence, regardless of intervening content. This possibility was tested in Experiment 5.

Even if the matrix subject is not maintained in active memory, and although the materials for Experiment 3 used temporal adjuncts instead of rational clauses headed by *in order*, it may have been the case that *before* and *after* still induced some sort of causal relation that caused participants to reactivate the subject of the preceding clause. However, this second possibility seems unlikely, since even the items with the connective *while*, which seems more neutral, showed no difference in the two conditions.

To reiterate, because Experiment 3 did not include a control condition with an unrelated verb, we cannot know whether the lack of difference between the two conditions represents equal facilitation or a lack thereof. There are several reasons why the similar N400s seen in the two conditions may have been the result of a lack of facilitation at the verb. First, there was an adverb intervening between the pronoun and the related verb. Thus, even if the pronoun activated semantic features of its antecedent, it may have been that this activation was short-lived, and by the time the verb was reached, it was not strong enough to cause a reduced N400.6

---

6See footnote 4 for why the adverb was included.
It also could have been the case that the pronoun was not being interpreted at all, or at least not right away. However, previous experiments (Arnold et al., 2000; Arnold & Lao, 2015) have shown that pronouns can be interpreted quite rapidly, and it is unclear why participants would not have also resolved the referent of the pronoun quickly in this experiment. Still, it could have been that the comprehension questions asked and the sentences themselves allowed for shallow interpretations without resolution of the pronoun (see Koornneef & Reuland, 2016).

Another possibility is that the pronoun was being interpreted immediately, but that this interpretation does not cause reactivation of the semantic features of the antecedent sufficiently to facilitate the processing of a semantically related verb. Although the relatedness of the antecedent noun and embedded verb was shown to induce a reduced N400 in Experiment 1, it could be that the pronoun did not activate the relevant features to cause the facilitation.

In sum, the results of this experiment showed a lack of difference in the N400 of an embedded verb related to the matrix subject, whether the embedded subject was a pronoun referring to the matrix subject, or whether it was an unrelated name. Because an unrelated verb control was not included, it is unclear whether this null effect is due to equal facilitation in the two conditions, as was seen in Experiment 1, or whether it reflects a lack of facilitation in either condition. If these results are due to equal facilitation in the two conditions, then they begin to answer some of the questions left by Experiment 1. First, Experiment 3 suggests that a lack of difference in the facilitation at the embedded verb in Experiment 1 was not due solely to extra material in the matrix clause priming the related verb (see (6)). Experiment 3 controlled for this possibility by ensuring that the content of the matrix clause was neutral. Because a lack of difference at the related embedded verb was still seen, it is likely that this was not the cause of the similar effect in Experiment 1. Similarly, Experiment 1’s results were likely not due to in order reactivating the matrix subject. Although it is possible that there was a causal relation implied in some of the materials of Experiment 3, it was not present in many of the items, and a similar lack of difference in the two (critical) conditions was seen.

**Experiment 4**

Experiment 4 tested whether the relatively small or lack of effect seen in Experiment 3 was due to equal prediction of the critical word in each condition. Using a cloze task on Mechanical Turk, similar to Experiment 2, items were created from the critical items in Experiment 3 up to and including the adverb prior to the critical word. The task took approximately 40 minutes to complete (paired with a separate unrelated cloze experiment), and participants (60 native speakers of American English) were compensated $4.
Results

The results of Experiment 4 are given in Figure 12. A chi squared contingency test indicated that the related verb was significantly more predicted in the pronoun condition than in the control condition (7.6% vs. 1.8%, $\chi^2 = 66.2, p < 0.001$).

Discussion

This cloze experiment demonstrates that the related verb was not equally predicted in the two conditions, suggesting that equal prediction cannot independently account for the lack of significant effect in Experiment 3. However, although the related verb was predicted significantly more in the pronoun condition, this cloze probability was itself rather small (less than 8%). This is likely due to the fact that in Experiment 3, and hence in Experiment 4 as well, the materials were specifically designed to reduce the amount of general contextual priming of the critical word. In achieving this goal, we also reduced the predictability of the related verb. Because the the related verb was generally not predictable in either condition, it is unsurprising that the difference in the N400 response would be small or inexistent, since predictability is such an important component of the N400 (Lau, Namyst, Fogel, & Delgado, 2016).

These EEG and cloze results leave several remaining possible explanations for the lack of effect in Ex-
periments 1 or 3. Specifically, do sentential subjects remain active in memory such that associates of the subject will always be primed? If so, how long does that activation last? Does that differ for objects, since objects are less likely than subjects to be referred to again? Or are all arguments of a sentence maintained active for some period of time (e.g. to the end of the sentence)? Experiment 5 attempted to address some of these questions.

**Experiment 5**

One possible explanation for the results of Experiments 1 and 3 is that in both cases, the intended antecedent for the pronoun/PRO was the subject of the matrix clause. It may be the case that there is a differential activation of subjects as compared to objects of a sentence, such that (main-clause) subjects are maintained in active memory throughout the reading of a sentence. If this is the case, then upon reaching a verb related to this subject in the previous experiments, equal facilitation would occur, regardless of intervening material (i.e. whether the subject of the embedded clause was coreferent with the matrix subject). Experiment 5 used self-paced reading to examine the processing of pronouns and the following verbs in an embedded temporal adjunct, manipulating whether the intended antecedent to the pronoun was the subject or the object of the matrix clause.

**Materials and methods**

This self-paced reading experiment was run on two different platforms. The first used the Linger software package (Doug Rohde, MIT), with participants completing the experiment in person in a lab at the University of Maryland. The second used the Ibex website (Alex Drummond: http://spellout.net/ibexfarm/). The two versions are referred to below as the Linger experiment and the Ibex experiment.

**Participants**

Participants in the Linger experiment were 32 undergraduates recruited at the University of Maryland, and received credit in an introductory linguistics course for their participation. There were 36 participants in the Ibex experiment, recruited through Amazon Mechanical Turk (https://www.mturk.com/). They received $4.50 for their participation. Four participants were excluded from the Ibex experiment for low accuracy on the comprehension questions (less than 80%), for a total of 32 Ibex participants. This resulted in a final dataset of 64 participants across the two platforms. Participants in both versions of the experiment were native speakers of English, gave informed consent, and were naïve to the purpose of the experiment.
<table>
<thead>
<tr>
<th>Cond.</th>
<th>Description</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Target_subj, pron_target, related</td>
<td>The soldier remembered Brianna with deep fondness while he fought in the war.</td>
</tr>
<tr>
<td>b</td>
<td>Target_subj, pron_other, related</td>
<td>The soldier remembered Brianna with deep fondness while she fought in the war.</td>
</tr>
<tr>
<td>c</td>
<td>Target_obj, pron_target, related</td>
<td>Brianna remembered the soldier with deep fondness while he fought in the war.</td>
</tr>
<tr>
<td>d</td>
<td>Target_obj, pron_other, related</td>
<td>Brianna remembered the soldier with deep fondness while she fought in the war.</td>
</tr>
<tr>
<td>e</td>
<td>Target_subj, pron_target, unrelated</td>
<td>The soldier remembered Brianna with deep fondness while he cured the sick little boy.</td>
</tr>
<tr>
<td>f</td>
<td>Target_subj, pron_other, unrelated</td>
<td>The soldier remembered Brianna with deep fondness while she cured the sick little boy.</td>
</tr>
<tr>
<td>g</td>
<td>Target_obj, pron_target, unrelated</td>
<td>Brianna remembered the soldier with deep fondness while he cured the sick little boy.</td>
</tr>
<tr>
<td>h</td>
<td>Target_obj, pron_other, unrelated</td>
<td>Brianna remembered the soldier with deep fondness while she cured the sick little boy.</td>
</tr>
</tbody>
</table>

Table 5: Experiment 5 sample items.

**Materials**

Sentences were created by combining a matrix clause with a finite temporal adjunct. The matrix clause contained two noun phrases: one with a determiner and a noun that is stereotypically viewed as describing an individual of a particular gender (e.g. *soldier*, stereotypically viewed as male, and *florist*, typically seen as female) which we call the *target NP*, and one containing a name of the opposite gender. The temporal adjunct was headed by *while* and contained a pronominal subject. In a $2 \times 2 \times 2$ design, the stimuli varied in whether the target NP was the subject or the object of the matrix clause, in whether the pronoun referred to the target NP or to the other name, and in whether the verb in the temporal adjunct was semantically related or unrelated to the target NP. A sample set of stimuli is given in Table 5.

It was important that priming of the embedded verb in conditions (c) and (d) was not increased compared to conditions (a) and (b) simply due to the target being more recent when in object position. For this reason, a three-word filler phrase was inserted between the object and the start of the temporal adjunct to reduce the chance of priming due to recency.

To create the unrelated items in (e-h), we replaced the temporal adjunct from one set of items with the
adjunct from another. For example, the temporal adjunct in (e-h) in Table 5 was taken from another set of items, where *cure* was related to a separate target NP: *The doctor thought about Nicole every other day while he cured the sick little boy.*

Forty-eight sets of eight items in these conditions were distributed across eight lists in a Latin square design. 144 filler sentences were also included, such that each participant read a total of 192 sentences. Each sentence was followed by a comprehension question targeting information in various regions of the sentences. This was intended to increase the likelihood of participants reading for deeper comprehension during the task. For a subset (16) of the items, this question directly targeted the referent of the pronoun in order to ensure that the intended interpretation was obtained (e.g. *Who fought in the war?* or *Who cured the sick little boy?* for the set of items in Table 5).

The materials were the same for the Linger and Ibex versions of the experiment.

**Procedure**

This experiment followed the same basic procedure used in McCourt et al. (2015). Sentences were displayed in a moving-window self-paced reading display. Each sentence initially appeared on a blank screen masked by white dashes with spacing intact. Participants revealed each word sequentially by pressing the space-bar. Words appeared in place of their respective dashes non-cumulatively on each press of the space-bar. The order of presentation of items was randomized for each participant. Participants were instructed to read the sentences at a natural pace, making sure that they understood what they read. The experiment started with five practice items to familiarize participants with the task. Each sentence was followed by a two-way multiple choice comprehension question.

Both Linger and Ibex versions of the experiment followed the same procedure, the only difference being the platform used.

**Data analysis**

Reading times above 2000 ms or below 50 ms were excluded, resulting in a loss of 0.46% of data in the critical regions. Reading times were then converted to a log scale, and statistical analysis was performed in four consecutive regions: the pronominal subject of the temporal adjunct, the following verb, and the two following words. Analysis was performed on the Ibex and Linger data separately, as well as on the combined data.

We used linear mixed-effects regression models to determine reliability of effects associated with the experimental factors. The fixed effects in the model were the factors *target location* (whether the target NP was the subject or the object of the matrix clause), *pronoun referent* (target or other name), *verb*
relatedness (whether the verb in the embedded clause was semantically associated with the target NP), and the interactions of those factors. In addition, participants and items were crossed, starting with random intercepts and slopes and removing one level of complexity until the model converged with correlations of less than 0.9 in random effects in all regions, following Baayen (2008) and Barr, Levy, Scheepers, and Tily (2013). The final convergent model included random intercepts for participants and items, but no random slopes.

Results

Reading times

Overall, participants in the Ibex version of the Experiment 5 had faster reading times than in the Linger version ($p < 0.001$; unlogged means were 304.9 ms for Ibex and 425.4 ms for Linger). Logged reading times for Experiment 5 are given in Figure 13. The regions of interest are outlined, beginning with the pronoun in the temporal adjunct, and extending to two words past the embedded verb. Figure 13(a) gives the reading times obtained in the Linger experiment, and Figure 13(b) the results from the Ibex experiment. The combined results from both experiments are in Figure 13(c).

In the Linger data, Region 5 showed marginal interactions between target location and pronoun referent ($p = 0.07$), and between target location, pronoun referent, and verb relatedness ($p = 0.07$). Region 7 had a marginal main effect of target location ($p = 0.08$) as well as a marginal interaction between target location and verb relatedness ($p = 0.06$). Region 8 had a significant main effect of verb relatedness ($p < 0.05$) and significant interactions between pronoun referent and verb relatedness ($p < 0.05$) and between target location, pronoun referent, and verb relatedness ($p < 0.05$). There was also a marginal interaction between target location and verb relatedness ($p = 0.08$).

The Ibex data had no significant effects in any region (all $p > 0.1$)

In the combined data, there was a marginal interaction between target location and verb relatedness in Region 6 ($p = 0.06$). In Region 8 there was a significant main effect of verb relatedness as as significant interactions between pronoun referent and verb relatedness, and between target location, pronoun referent, and verb relatedness (all $p < 0.05$). There was also a marginal interaction between target location and verb relatedness ($p = 0.07$). There were no other effects in the combined data ($p > 0.1$ for all other effects in each region).
Figure 13: Experiment 5: Reading times in both versions of Experiment 5 (Linger (a), Ibex in (b)), as well as for the combined data (c).
<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Target_subj, pron_target, related</td>
<td>0.945</td>
</tr>
<tr>
<td>b</td>
<td>Target_subj, pron_other, related</td>
<td>0.773</td>
</tr>
<tr>
<td>c</td>
<td>Target_obj, pron_target, related</td>
<td>0.922</td>
</tr>
<tr>
<td>d</td>
<td>Target_obj, pron_other, related</td>
<td>0.773</td>
</tr>
<tr>
<td>e</td>
<td>Target_subj, pron_target, unrelated</td>
<td>0.914</td>
</tr>
<tr>
<td>f</td>
<td>Target_subj, pron_other, unrelated</td>
<td>0.914</td>
</tr>
<tr>
<td>g</td>
<td>Target_obj, pron_target, unrelated</td>
<td>0.859</td>
</tr>
<tr>
<td>h</td>
<td>Target_obj, pron_other, unrelated</td>
<td>0.898</td>
</tr>
</tbody>
</table>

Table 6: Accuracy on Experiment 5 comprehension questions asking about pronoun by condition.

Comprehension questions

Accuracy on the subset of questions targeting pronoun interpretation by condition are given in Table 6. We compared the mean accuracy across conditions using a linear mixed-effects model with participants and items as random effects. The model found a significant main effect of pronoun referent on question accuracy ($p < 0.001$) as well as a significant interaction between pronoun referent and verb relatedness ($p < 0.01$). Pairwise analyses revealed that these effects were driven by conditions (b) and (d), which had significantly lower accuracy than other conditions.

Discussion

Reading times

This experiment tested the following hypotheses.

1. Simple priming/predictability causes facilitation of related verbs in the adjunct compared to unrelated verbs.

2. There is a general preference for the embedded pronoun to refer to the subject of the matrix clause that will be reflected in processing.

3. The semantic content of the main clause subject, but not the object, remains active throughout the sentence.

4. Objects are less important for future discourse, and their content is not held active, but a pronoun may reactivate its content.

5. Both 3 and 4 may be true. The semantic content of a main clause subject may remain active throughout the sentence, and pronouns may reactivate the semantic content of their antecedents.
If Hypothesis 1 were true, the verb *fought* would be facilitated simply due to the presence in the sentence of the related noun *soldier*. If this were the case, then conditions (a-d) should show facilitation in processing at or after the embedded verb. In other words, (a-d) should have faster reading times than (e-h) in the critical region.

There may have been a small priming effect in the Linger versions of the experiment, but only when the target noun was in the object position of the matrix clause. Conditions (c) and (d) had faster reading times at the verb and in the following region than conditions (a) and (b) and conditions (g) and (h). This effect was small, though. In the region following the verb, there was a marginal main effect of target location as well as a marginal interaction of target location and verb relatedness. This suggests that when the target noun was closer to the related verb, the verb may have been easier to process. It is possible that a bigger effect was not seen due to the three-word buffer included before the temporal adjunct. The material intervening between the verb and the target noun may have reduced a priming effect. And when even more material intervened, i.e. when the target was in subject position, facilitation due to priming was eliminated.

Although there may have been priming of the verb due to a recent related target noun in the Linger version of the experiment, no priming effects were seen in the Ibex data. In the combined data, there was no overall effect of relatedness of the verb in regions 5-7. There is a marginal interaction of target location and verb relatedness at the verb, consistent with a priming recency effect. In addition, there is a significant main effect of verb relatedness in region 8, i.e. two words after the embedded verb. However, this main effect is driven by a single condition. Besides condition (e), all other conditions overlap in this region, and there are significant interactions between pronoun referent and verb relatedness as well as between pronoun referent, verb relatedness, and target location. Because it is not the case that all conditions with an unrelated verb were slower than conditions with a related verb, there is no evidence for simple priming of verbs due to the presence of a related target noun present somewhere previously in the sentence.

Hypothesis 2 predicts that processing of the pronoun may be facilitated when it refers to the subject rather than the object, leading to faster reading times at or after the pronoun in (a,d,e,h) compared to (b,c,f,g). If this were the case, we would expect an interaction between pronoun referent and target location. This interaction was not significant in any region in either of the experiments or in the combined data, although it was marginal at the pronoun in the Linger version of the experiment, which indicates that there may be an advantage for pronouns referring to the matrix subject rather than the object. But whatever advantage there may be is small and was not exhibited in the Ibex version of the experiment.

Hypothesis 3 predicts facilitation of a verb that is related to the subject, leading to faster reading times when the target noun is in subject position and is semantically related to the embedded verb. If this were the case, conditions (a-b) should have faster reading times than other conditions at or after the embedded
verb, which was not the case in either version of the experiment. Indeed, as noted earlier, conditions (c-d) were faster than (a-b) at the verb and in the following region for the Linger experiment. In the Ibex data, there was no difference in reading times across conditions. Therefore we saw no evidence for the semantic content of the subject remaining active throughout the sentence.

Under Hypothesis 4, when the pronoun in the embedded clause refers to the matrix object, there should be facilitation of a related verb compared to when the pronoun refers to an unrelated noun or when the verb is unrelated. In this case, (c) should have faster reading times at or after the embedded verb than (d) and (g). Similarly, when the pronoun refers to the subject of the matrix clause, it may reactivate its content, leading to faster reading times in (a) than in (b) or (e). Although condition (c) did show faster reading times than (g) at the verb and in the following region, this appears to be due to the simple priming and recency effect discussed above, as there was no difference between (c) and (d), or between (a,b,e). There was no evidence for the pronoun reactivating the semantic content of its antecedent such that reading times would be reduced.

Hypothesis 5 predicts faster reading times at or after the embedded verb for (a-c) than for (d-h). For (a-b), the content of the target would remain active. Furthermore, for (a) and (c), the semantic content of the target would be reactivated, leading to facilitation of the related verb. This was not seen.

Although Experiment 5 did not provide support for any of these hypotheses, it also does not provide strong evidence against them. It may be, for example, that subjects are privileged and maintained in active memory more than objects, or that pronouns reactivate the semantic content of their antecedents, but that this (re)activation does not lead to faster reading times for later related content.

The only significant effect seen was one that the design did not directly test for. In region 8, two words after the embedded verb, there was a significant main effect of verb relatedness, as well as significant 2- and 3-way interactions between the pronoun referent and verb relatedness and between target location, pronoun referent, and verb relatedness. These effects were also present in the Linger data alone, but were not significant in the Ibex data. The effects were driven by a single condition (e), which had slower reading times than all other conditions, and this was true numerically even in the Ibex data. This may be the result of a semantic anomaly affect that was not controlled for. It may be that in that condition, the sentence became increasingly implausible. In one set of items, represented in (7), although the verb arrange in condition (e) is not as semantically related to the target cop as the verb arrest in condition (a), it is plausible that a cop would arrange things. However, it is much less plausible that the cop would be arranging flowers.

(7)  

a. The cop offended Rachel with bad language while he arrested a thief in the shop.

e. The cop offended Rachel with bad language while he arranged some flowers into a beautiful
bouquet.

g. Rachel offended the cop with bad language while he arranged some flowers into a beautiful bouquet.

Items such as this may have surprised participants in this region, leading to slower reading times. However, it is not clear why this would not also occur in condition (g), in which the cop is also described as arranging flowers. It could be that the matrix event is contributing to people’s expectations of what is plausible in the embedded clause. A cop may be even less likely to arrange flowers while offending someone with bad language, but it may be less surprising that he would do so while someone was offending him.

This may have also contributed to the lack of effects seen in the critical comparisons. The materials were designed so that the verb would be somewhat plausible in either the related or unrelated conditions. And although the related verb is more closely associated with the target noun, it could be that this difference in association was not strong enough to cause reading time differences.

Comprehension questions

One interesting and unexpected result came in the form of accuracy on the subset of comprehension questions directly targeting the referent of the pronoun. Conditions (b) and (d) had significantly lower accuracy on pronoun interpretation, and participants were more likely to respond with the target noun when asked about the related verb in these conditions. This either means that participants ignored gender biases and interpreted the pronoun as the target, or that they initially interpreted the pronoun according to gender biases, but misremembered the sentence in answering the comprehension question. The former would help explain why effects of verb relatedness were not stronger in reading time measures. In either case, results from the comprehension questions show that participants were associating the related verb with the target noun, sometimes even more strongly than intended.

General discussion

This paper attempted to provide evidence on the interpretation of PRO and pronouns. The experiments were unable to find definitive evidence in the N400 component or in self-paced reading times about the timecourse of interpretation of the anaphora, or in fact whether the anaphora were resolved at all. Although small effects were seen in the N400, these only became evident when ERPs were baselined unconventionally in the window 0–100 ms post stimulus onset. Although there was some justification for this, as the control condition in each experiment introduced a new discourse referent, which is known to cause a large negative
response (Burkhardt & Roehm, 2007), this effect was not sustained to the critical word, and it is unclear that re-baselining the data was justified. And in either case, the small N400 differences seen were not statistically significant, even after re-baselining.

Previous literature suggests that pronouns and PRO can be interpreted quite rapidly (Arnold et al., 2000; Kwon & Sturt, 2015; Green et al., in preparation). Therefore, it seems unlikely that the anaphora was not being resolved in these experiments, or that it took longer than in previous experiments, and the lack of relevant facilitation effects was not likely due lack of interpreting the pronoun/PRO. Instead, it is more likely that the resolution of the anaphora either does not cause rapid reactivation of the semantic content of the antecedent in memory, or that such reactivation is not sufficient to cause facilitation (in the form of a reduced N400 or faster reading times) of semantically related following material beyond facilitation caused by other factors such as simple priming. This is true regardless of whether the antecedent is the subject or the object of the matrix clause, so we found no evidence of preferential treatment in memory for subjects over objects.

There are at least two interesting implications from these experiments. First, Experiments 1–4 showed a discrepancy between cloze results and N400 responses. Although there was a difference in cloze probability of the related verb in the PRO and control conditions, this difference was not reflected in the N400 response. This may provide additional evidence that the N400 does not simply track cloze probability (see Kutas and Federmeier (2011) for discussion). However, the cloze probability of the critical verb in Experiments 2 and 4 were relatively small in all conditions. Because of this, it may be unsurprising that there would only be only a small, even perhaps undetectable, reduction of the N400. Second, the comprehension question results from Experiment 5 have implications for studies attempting to use gender biases to manipulate interpretation (e.g. Parker et al. (2015)). Participants in this experiment were more likely to ignore the gender bias in resolving the pronoun when it was the subject of a verb related to an antecedent stereotypically of the opposite gender from the pronoun.

One other possible explanation for the lack of effects seen is that all referents are maintained equally active in memory for the duration of the sentence, despite possible differences in their discourse status. Or, it could be that subjects really do have preferential status in memory, and that this caused the results in Experiment 1, but that self-paced reading times are not sensitive to this in the same way that the N400 is, causing the lack of priming seen in Experiment 5. Or it could be that the result in Experiment 1 was due to simple contextual priming not reflected in cloze probability, as was seen in Kutas and Federmeier (2000), and discussed in Ettinger et al. (2016).
References


Green, J. J. (submitted). Non-subject control of temporal adjuncts.


