Lingering Misinterpretations of Garden Path Sentences Arise from Competing Syntactic Representations

Timothy J. Slattery

University of South Alabama

Patrick Sturt

University of Edinburgh

Kiel Christianson

University of Illinois, Urbana-Champaign

Masaya Yoshida

Northwestern University

Fernanda Ferreira

University of South Carolina

Correspondence to:

Timothy J. Slattery

Psychology Department

University of South Alabama

Email: slattery@southalabama.edu

Word count: 8885

Running Head: Lingering Misinterpretations

Keywords: Eye Movements; Reading; Good Enough Sentence Processing; Parsing; Reanalysis

ABSTRACT

Recent work has suggested that readers' initial and incorrect interpretation of temporarily ambiguous ("garden path") sentences (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001) sometimes lingers even after attempts at reanalysis. These lingering effects have been attributed to incomplete reanalysis. In two eye tracking experiments, we distinguish between two types of incompleteness: the language comprehension system might not build a faithful syntactic structure, or it might not fully erase the structure built during an initial misparse. The first experiment used reflexive binding and the Gender Mismatch paradigm to show that a complete and faithful structure is built following processing of the garden-path. The second experiment used two-sentence texts to examine the extent to which the garden-path meaning from the first sentence interferes with reading of the second. Together, the results indicate that misinterpretation effects are attributable not to failure in building a proper structure, but rather to failure in cleaning up all remnants of earlier attempts to build that syntactic representation.

Introduction

Sentences containing temporary local ambiguities such as the "garden-path sentence" in (1) have been exploited by psycholinguists for decades (Ferreira & Henderson, 1991; Frazier & Rayner, 1982) as a way to reveal the mechanisms of language comprehension. Garden-path sentences reveal the parser's preferences for resolving syntactic ambiguities and for recovery when incorrect syntactic decisions are initially made.

(1) While Anna dressed the baby that was cute and cuddly played in the crib.

A central assumption in the study of language comprehension has been that sentence meaning is derived from a complete structural representation built up from the component parts of a sentence into a fully specified syntactic structure; if no such structure is obtained, then only a coarse interpretation of the input sentence will be generated. MacDonald, Pearlmutter, and Seidenberg's (1994) comment that there might be situations in which "the communicative goals of the listener can be achieved with only a partial analysis of a sentence, but we view these as degenerate cases" (p. 686) exemplifies this assumption.

Only recently have there been studies examining the post-repair representations that comprehenders build for such temporarily ambiguous sentences (Christianson, et al., 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, 2003; Ferreira, Christianson, & Hollingworth, 2001; Swets, Desmet, Clifton, & Ferreira, 2008). In one such study, Christianson et al. (2001) examined the traditional assumption that full resolution of a local syntactic ambiguity is necessarily performed as part of the process of deriving a robust interpretation of a sentence. Christianson et al. asked participants to read sentences containing ambiguities as in (1) and then answer comprehension questions as in (2).

(2) Did Anna dress the baby?

Christianson et al. reasoned that, if readers construct the interpretation that is compatible with the global structure of the sentence, the answer to this question must be "No" as *the baby* is the subject of the matrix clause but not the object of the embedded clause, whereas the answer to a related question (Christianson et al., 2006) exemplified in (3) must be "Yes."

(3) Did Anna dress herself?

Christianson et al. (2001, 2006) demonstrated in a number of experiments, however, that syntactic manipulations of the garden path, including clause order, disambiguation, and length of ambiguous region, affected accuracy rates on follow-up comprehension questions. None of these manipulations altered the lexical content of the sentences, nor the inferences likely to be drawn from those collections of lexical items. Moreover, Christianson et al. (2006) found that the likelihood of answering questions such as (3) correctly was related to readers' working memory capacity. The authors therefore argued that full reanalysis required more cognitive effort than readers with lower working memories were willing or able to expend, and so these readers settled for “Good Enough” reanalysis and interpretations (Ferreira, Bailey, & Ferraro, 2002; Ferreira et al., 2001; Ferreira & Patson, 2007).

Several recent studies have corroborated Christianson et al.'s observation that misanalyses appear to linger (Kaschak & Glenberg, 2004; Staub, 2007a; Sturt, 2007; van Gompel, Pickering, Pearson, & Jacob, 2006), though researchers differ on why this might be. For instance, Christianson et al. (2001; see also Christianson et al., 2006) attributed the lingering effect to a failure to fully reanalyze the partial, and ultimately incorrect, syntactic structure constructed during the initial parse of the garden path sentence. However, Kaschak and Glenberg (2004) suggested that the persistence of previous syntactic structure stems from episodic memory traces of the initial structure, whereas Sturt (2007) suggested that the effects are due to semantic persistence. Van Gompel et al. (2006) observed that the structure of initial misanalyses of the sort of garden-path sentences used in Christianson et al. (2001) primed subsequent production, but remained agnostic as to whether the source of this structural priming was episodic memory or incomplete syntactic reanalysis.

Under traditional models of parsing – serial (Frazier & Rayner, 1982), partially parallel

(Gibson, 1991; 2000), fully parallel (MacDonald et al., 1994), or pragmatically-mediated parallel (Crain & Steedman, 1985) – two ways of handling temporary ambiguities exist. In the first, the ambiguous choice point in the parse is detected (either as a misparse, or as an alternative disfavoured parse), and reanalysis occurs, bringing the entire structure into compliance with the grammar and generating the correct semantic interpretation for the string. It is possible that a previously top-ranked but subsequently devalued structure in a ranked-parallel system (e.g., MacDonald et al., 1994) could remain partially activated; however, we are not aware of any work examining effects of such re-ranking on later processing. The second possibility is that the ambiguity is not noticed at all. This could happen if, at the ambiguous choice point, the parser initially settled on the eventually correct interpretation and thereby avoided the need for reprocessing. Alternatively, the incorrect parse and interpretation might be chosen, but the disambiguating information might not trigger reanalysis. In either case of insensitivity to the ambiguity, one would not expect to observe the classic eye movement signature of syntactic reanalysis, namely long fixation times on the disambiguating region, often accompanied by regressive eye movements from the disambiguating word and re-reading of the ambiguous text (Frazier & Rayner, 1982).

Thus, the traditional view of how such ambiguities are processed does not predict any lingering effects of the ambiguous structure on the processing of subsequent material, either because the initial parse was fully corrected prior to moving on, or because the ambiguity was entirely ignored. (But cf. Levy, Bicknell, Slattery, and Rayner (2009), who describe a computational account of lingering uncertainty in sentence processing. We return to Levy et al.'s proposal in the General Discussion.) If, however, processing at either the syntactic or semantic level (or both) is incomplete (or incorrect) prior to moving on to subsequent material, then one would expect lingering effects of the temporary ambiguity. Under a Good Enough view of sentence processing (Christianson et al., 2001; Ferreira et al., 2001, 2002; Ferreira & Patson, 2007), this is precisely the prediction. The Good Enough (GE) approach to language comprehension assumes that the goal of a comprehender is not to create a wholly accurate representation of the input, but rather to construct a sensible meaning for this input quickly and efficiently. This approach allows for the possibility that comprehenders create interpretations of the input that seem “good enough” for them to continue with the task of interpreting new input, but upon closer examination (often taking the form of comprehension probes) it becomes apparent that they have lingering misinterpretations of this input. These misinterpretations persist despite evidence that reanalysis is performed and new interpretations are generated. However, it is not yet known why the misinterpretation lingers past disambiguating material. Here we will consider two alternative conceptions of the reanalysis process that are capable of accounting for these lingering misinterpretations. The first possibility is that during reanalysis, the parser attempts to revise the original syntactic structure, but this occasionally yields a structure that is not fully detailed or well-formed. In this view, the syntactic representation itself is incomplete, disconnected, or just plain wrong, which then undermines the quality of the semantic interpretation built from it. As a consequence, this view would predict a strong lingering influence of the structurally inadequate representation, resulting in a global slowdown in reading until input arrives that allows for a stable interpretation. Such a global slowdown would be analogous to what is observed in real-world instances of increased cognitive load due to grammatically illicit input, for example, while processing speech of non-native speakers (Gass & Varonis, 1984; Munro & Derwing, 1995).

The second possibility is that the parser initially creates an incorrect parse for the ambiguous material, encounters the disambiguating word, and then, during reanalysis, builds a new structure that is complete, fully specified, and faithful to the input, but does not completely prune the original mis-analysis. The parser then moves on to process new input on the basis of the newer (correct) syntactic representation. As the overall representation is complete, and capable of yielding a stable interpretation, this view does not predict a global slow-down in reading following reanalysis. However, the original un-pruned analysis may lead to lingering misinterpretations, which can influence processing in certain circumstances. According to this hypothesis, because the parser is working on the basis of a correct representation, it can move on to new input. Thus, on the former view, it is the structure that is incomplete, and merely "good enough," whereas on the latter view, it is the process of reanalysis that is merely "good enough."

We explored these two possibilities with two eye-tracking experiments in which reading patterns on text subsequent to the point of disambiguation were analyzed to determine the extent of reanalysis. The first experiment focused on whether a well-known grammatical constraint is obeyed following presumed reanalysis; the second experiment examined two-sentence sequences in which a garden-path sentence was followed by a sentence that assumed the interpretation associated with full reanalysis. Together, these two experiments allow us to assess the completeness of the syntactic representation built after reanalysis, and to determine whether the representations built prior to reanalysis linger and affect later on-line processing.

For Experiment 1, to investigate the completeness of the syntactic representation that the parser constructs after a garden-path, we took advantage of a well-established generalization regarding the processing of reflexives pronouns. Previous studies on the processing of reflexives have demonstrated that an antecedent is searched for immediately upon encountering a reflexive in the input, and a disruption is seen if the gender of the reflexive mismatches that of its antecedent (the so-called Gender Mismatch Effects (GMME); Sturt, 2003). Furthermore, existing studies that tracked eye movements during reading have shown that this antecedent search process respects grammatical constraints on coreference, at least in results for early (first-pass) eye-movement measures. Thus, the reading slowdown associated with the GMME is observed as long as the link between the reflexive and its antecedent is grammatically permitted (Sturt 2003). In the theoretical literature, it has been established that the distribution of noun phrases is governed by the structural conditions known as Binding Principles (Chomsky 1981, among others). It has been established that three types of noun phrases - reflexives, non-reflexive pronouns and names - obey three different structural conditions.

Reflexives, the focus of Experiment 1, are constrained by Binding Condition A. For example, an NP in the possessor position as in (4a) is not a legitimate antecedent for the reflexive pronoun, unlike an NP in the Nominative (subject) position as in (4b).

(4) a. *David's mother gave himself approximately five days to reply.

b. Dana's father gave himself approximately five days to reply.

This contrast is captured by the structural differences between possessive and nominative NPs: in (4), the nominative NP c-commands the reflexive in its local domain, but the possessive does not (Chomsky 1981; Reinhart & Reuland 1993). Processing studies have demonstrated the GMME for grammatically licit antecedents of reflexives, while an equivalent mismatch effect is not observed—or is observed only after some delay— for grammatically illicit antecedents of reflexives (Sturt, 2003), although such non-grammatical antecedents can cause competition effects (Badecker & Straub, 2002; Runner et al., 2003, 2006).

Given that the initial processing of reflexives respects structural constraints, and given that the GMME is seen only when the reflexive can be linked to its antecedent grammatically, we can use this effect as a probe for the structure constructed by the parser. If we observe a GMME associated with the processing of a reflexive pronoun, it would indicate that the reflexive and its antecedent are in a specific syntactic relation (the antecedent c-commands the reflexive, or the antecedent is a co-argument of the reflexive), and it further would suggest that the parser constructs such a detailed structural representation. In the context of the processing of garden path sentences, if the parser is led down a garden path, and as a result, fails to construct the globally correct syntactic structure after reanalysis, such a structure would not support the grammatical relation between the reflexive and its antecedent. If this is the case, we expect not to observe a GMME after the parser is garden-pathed. On the other hand, if a proper syntactic representation is completed during reanalysis, we would expect to observe the GMME regardless of temporary ambiguity. Moreover, if the globally correct syntactic structure has not been computed by the time the reflexive is reached, then we should see evidence of processing difficulty due to the ambiguity. This is because the reflexive triggers a search for its antecedent and the parser tries to link the two; yet this attempt will fail because, regardless of gender matching, there will be no accessible antecedent for the reflexive. Thus, in this case, we expect a slowdown related to the ambiguity but no effect of reflexives. In this way, we can diagnose whether syntactic reanalysis is completed by making use of the GMME and its sensitivity to the structural binding constraint, and we can look for overall evidence of difficulty for temporarily ambiguous sentences, relative to appropriate baselines.

Experiment 1

Participants read sentences such as those in (5) while their eye movements were monitored. We varied reflexive gender matching (match vs. mismatch) and ambiguity (garden path vs. non-garden path). Ambiguity was manipulated by the presence or absence of a comma after the embedded verb. The result is a 2×2 factorial design. In the garden path conditions, the first clause (preposed adverbial adjunct clauses) consisted of a subordinate-clause ambiguous syntactic structure. Therefore, at the point of the embedded verb *telephoned*, the structure was ambiguous in terms of whether the verb was transitive or intransitive. In the non-garden path conditions, the structure was unambiguous due to the presence of the comma (Christianson et al., 2001). Because the comma was inserted between the verb and the subsequent NP, the embedded verb could only be interpreted as intransitive.

(5) Sample set Experiment 1 stimuli[1]

[1] Note that our design manipulates definitional gender (e.g. *mother* and *father* are female and male by definition), as opposed to stereotypical gender (e.g. *nurse*, *surgeon*) used in some other studies such as Sturt (2003). Previous work has

a. *Garden Path/Match*

After the bank manager telephoned David's <u>father</u> grew worried and gave himself approximately five days to reply.

b. *Garden Path/Mismatch*

After the bank manager telephoned David's <u>mother</u> grew worried and gave himself approximately five days to reply.

c. *Non-Garden Path/Match*

After the bank manager telephoned, David's <u>father</u> grew worried and gave himself approximately five days to reply.

d. *Non-Garden Path/Mismatch*

After the bank manager telephoned, David's <u>mother</u> grew worried and gave himself approximately five days to reply.

The sentences were divided into regions for the purpose of data analysis, as shown in Table:

Table 1

With this design, we predict the following. First, consistent with previous studies, we should observe a main effect of structure, such that the garden path sentences will show signs of processing difficulty, including inflated reading times on the disambiguating verb (e.g., *grew*), compared to the non-garden path (comma-control) conditions. Second, we predict slow reading times on the reflexive (*himself*) in the mismatch conditions relative to the match conditions, as previously observed.

---

shown that the gender mismatch cost is greater for definitional than stereotypical mismatches in some measures, although the overall pattern is similar (see Kreiner, Sturt & Garrod, 2008, Experiment 1).

Most importantly, if the parser successfully constructs a detailed syntactic structure after initial misanalysis, then this slow-down should be observed in both the garden-path and the non-garden-path conditions, resulting in a main effect of gender matching that should not interact with ambiguity. On the other hand, if the parser fails to construct a coherent syntactic structure after initial misanalysis, there are several possible consequences. First, if *David's father/mother* is not analyzed as the matrix subject at this point, then the reflexive does not have an accessible antecedent. This lack of an accessible antecedent should lead to increased reading times on the reflexive in the ambiguous conditions relative to the unambiguous conditions, over and above any residual difficulty due to spill-over from the garden path.

Second, as neither *David* nor *father/mother* would be possible as a grammatical antecedent, there should be little reason to choose one or the other as the antecedent. Therefore, the GMME should be absent in the ambiguous conditions, or at least weaker than in the unambiguous conditions, where *mother/father* is the only grammatically possible antecedent, leading to a strong GMME penalty in the non-garden path/mismatch condition relative to the non-garden path/match condition. Thus, this account would predict a main effect of ambiguity at the reflexive region, with ambiguous conditions being read more slowly than unambiguous conditions, with a possible interaction of gender matching with ambiguity as well.

A third possibility is that the parser does construct a detailed syntactic analysis during reanalysis, but only on a portion of the trials. This may occur, for example, in a model where parses are constructed probabilistically, picking up on cues in the input that favor or disfavor the likelihood of constructing particular structures (see, e.g. Lewis & Vasishth, 2005). Thus when readers encounter the reflexive, it would result in a mixture of trials in which they have a detailed syntactic analysis allowing for a GMME, and trials without the detailed syntactic analysis required for a GMME. This would predict a similar data pattern to the above possibility; an interaction between gender matching and ambiguity with a smaller GMME for the ambiguous conditions, and a main effect of ambiguity with slower reading in the ambiguous conditions.

Thus, the claim that the parser fails to create a detailed syntactic representation following reanalysis would be supported by an interaction at the reflexive region. The relative difficulty of the mismatch condition (compared with its match control) would be present only for the unambiguous non-garden path conditions, and absent, or much weaker for the ambiguous garden path conditions (i.e. if the parser were to fail to create the detailed syntactic representation but only on a proportion of the trials),. In contrast, if a detailed representation is constructed and utilized by the parser, we expect a main effect of gender matching, with a statistically equivalent cost for the mismatch condition, regardless of whether the sentence caused a garden path.

Method

*Participants***.** Twenty-four undergraduate students from the University of Edinburgh participated in the experiment for course credits (a further four participants were run, but the data could not be used because of calibration problems).

*Apparatus***.** Eye-movements were recorded via an SR Research Ltd. Eyelink 1000 eye-tracker which records the position of the reader’s eye twice every millisecond, and has a high spatial resolution of 0.01°. Subjects were seated 81 cm away from a 19-inch ViewSonic VX922 CRT monitor. Head movements were minimized with chin and head rests. Eye movements were recorded from the right eye. Texts were presented in 18pt Times Roman black font on a white background.

*Materials***.** Thirty-two items like those in (5) were prepared. The main clause contained the reflexive pronoun, and it had the same structure across conditions. The main clause always had a coordinated VP with two conjuncts. The reflexive always appeared in the second conjunct. We wrote the sentences in this way because the verb phrase in the first conjunct is the locus of reanalysis, and therefore, the reading time slowdown caused by the reanalysis could spill over to the object position of the verb, and could mask the reading pattern on the reflexive. We used the masculine reflexive *himself* in half of the items and the feminine reflexive *herself* in the other half of the items. For any given item, the reflexive was the same in all four conditions. Furthermore, the gender of the head noun of the main clause subject was manipulated. We used kinship terms for the main clause subjects, which matched in gender of reflexive in match conditions, and mismatched in mismatch conditions. Finally, the gender of the possessor noun in the main clause subject (which was always a male or female name) always matched the gender of the reflexive. Thus, the possessor noun was identical across four conditions of any given item, but the head noun differed between the match and mismatch conditions. The matrix subject head nouns of the match and mismatch conditions did not differ in log frequency of occurrence in the 90 million word written section of the British National Corpus (Match: ln(freq) = 8.07; Mismatch: ln(freq) = 8.06; $t < 1$). Moreover, the length of the whole matrix subject region was matched on number of characters (Match: 16.19 chars; Mismatch: 16.44 chars; $t < 1$).

*Procedure***.** Each trial began with a gaze trigger, which consisted of a black square presented in the position of the first character of the text. Once a stable fixation had been detected on the gaze trigger, the sentence was presented in full. The participant pressed a button on the button box to indicate that he/she had finished reading the sentence. At this point, the sentence disappeared, and, in 50% of the trials, a yes/no comprehension question was presented, which the participants answered by pressing the appropriate button on the button box. The comprehension questions did not probe the antecedent of the reflexive or the interpretation of the subordinate clause ambiguity. Sentences were presented in a random order intermixed with 78 filler sentences of varying structures, all of which were grammatical, including twelve that had the subordinate clause-main clause ordering (without garden paths, as the verbs were used transitively).

Results

Prior to analysis, fixations shorter than 80 ms (2.6%), trials with fixations longer than 1200 ms (.04%), and data points that were three standard deviations above the condition mean (< 2%)

were removed from the data record. For data analysis purposes, sentences were divided into critical regions, as shown in Table 1. For each region, 2 (ambiguity) x 2 (reflexive gender) x 4 (list) (Pollatsek & Well, 1995) mixed design ANOVAs were performed, with participants (F_1) and items (F_2) as random effects. Reading measures analyzed were first-pass time, go-past time (also called regression path duration, the cumulated fixations from first entering the region until leaving it to the right, including regressive fixations out of the region), and total time. For first pass and go-past, we excluded from analysis the trials in which the first fixation in the region occurred after the reader had already fixated material to the right of the region (i.e. skips). For total time, we excluded trials in which there were no fixations at all in the region. These excluded cases did not contribute to the relevant cell mean.

Significant results reported below were obtained at or below p = .05. The means of the various eye movement measures are given in Table 2. Though the most pertinent data are those of the critical region in sentence 2, we will present the reading data by region of interest in the order in which the regions are normally read.

Table 2

Connective Plus Subordinate Subject Region: The first pass time in this region was marginally longer in the garden path condition than in the non-garden path condition, $F_1(1, 20) = 3.02$, $MSe = 13464.7$, $p = 0.098$; $F_2(1, 28) = 3.60$, $MSe = 15800.3$, $p = 0.068$. However, there was no main effect of gender match, Fs < 1, and no significant interaction between ambiguity and gender,$F_1(1, 20) = 1.26$, $MSe = 19703.7$, $p = 0.275$; $F_2(1, 28) = 2.18$, $MSe = 14903.7$, $p = 0.151$. Note that go-past time for this region is identical to first-pass time by definition.

In total time, there was a significant main effect of ambiguity, $F_1(1, 20) = 31.54$, MSe =

54109.1, $p < 0.001$; $F_2(1, 28) = 37.02$, $MSe = 64696.6$, $p < 0.001$, as this region was fixated for more overall time in the garden-path condition than in the non-garden-path condition. There was no main effect of gender (*F*s< 1) and no interaction (*F*s < 1).

Subordinate Verb Plus Matrix Subject Region: There were no significant effects in first pass in this region (all *F*s < 1). In the go-past time measure, there was a significant effect of ambiguity, $F_1(1, 20) = 10.70$, $MSe = 27010.2$, $p = 0.004$; $F_2(1, 28) = 10.88$, $MSe = 33742.1$, $p = 0.003$, as the garden path conditions yielded longer times than the non-garden path conditions. Given that disambiguation was in the next region, this could be indicative of a preview effect or the result of mislocated fixations (Drieghe, Rayner, & Pollatsek, 2008). However, there was also a comma in this region for the non-garden path stimuli that wasn't there in the garden path stimuli (see Staub, 2007b). There was no main effect of gender and no interaction (all *F*s < 1)

As with the first region, there was a main effect of ambiguity on the total time measure of the subordinate verb plus matrix subject region, $F_1(1,20) = 37.34$, $MSe = 53542.7$, $p = 0.001$; $F_2(1, 28) = 42.11$, $MSe = 67926.7$, $p < 0.001$, as readers spent more time fixating on this region in the garden path conditions than in the non-garden path conditions. Additionally, there was a main effect of gender, $F_1(1, 20) = 25.57$, $MSe = 35314.0$, $p < 0.001$; $F_2(1, 28) = 14.38$, $MSe = 75028.5$, $p = 0.001$, as match conditions were read faster than the mismatch conditions. However, there was no significant interaction between gender and ambiguity (all *p*s > .05)

Disambiguation Region: This region showed robust evidence of a garden path effect, with ambiguous conditions being read more slowly than unambiguous conditions in first pass time, $F_1(1, 20) = 12.83$, $MSe = 1838.5$, $p = 0.002$; $F_2(1, 28) = 12.22$, $MSe = 2628.6$, $p = 0.002$, go-past time, $F_1(1, 20) = 31.70$, $MSe = 35963.3$, $p < 0.0001$; $F_2(1, 28) = 27.77$, $MSe = 57201.7$, $p < 0.001$, and total time, $F_1(1, 20) = 18.68$, $MSe = 11630.9$, $p < 0.001$; $F_2(1, 28) = 18.32$, $MSe = 15748.1$, $p = 0.001$. The main effect of gender was not significant in first pass or go-past time, Fs < 1. However, there was a gender effect in total time, $F_1(1, 20) = 8.66$, $MSe = 8701.0$, $p = 0.008$; $F_2(1, 28) = 7.40$, *MSe* = 11466.5, $p = 0.011$, as total time was shorter for the match than for the mismatch conditions. Finally, there was no indication of an interaction between gender and ambiguity in any of the measures, *F*s < 1.

Spill-over Region: As with the disambiguation region, there was a significant effect of ambiguity in all three measures analyzed: first pass, $F_1(1, 20) = 6.96$, *MSe* = 2952.1, $p = 0.016$; $F_2(1, 28) = 10.45$, *MSe* = 2804.0, $p = 0.003$, go-past, $F_1(1, 20) = 64.76$, *MSe* = 8368.4, $p < 0.0001$; $F_2(1, 28) = 26.90$, *MSe* = 23199.0, $p < 0.0001$, and total time, $F_1(1, 20) = 16.11$, *MSe* = 8706.1, $p = 0.001$; $F_2(1, 28) = 12.56$, *MSe* = 15482.5, $p = 0.001$. The GMME effect was absent in first pass and go-past time, *F*s < 1. However, in total time, which includes regressions from subsequent regions, there was a significant gender effect, $F_1(1, 20) = 25.18$, *MSe* = 7287.5, $p < 0.001$; $F_2(1, 28) = 13.78$, *MSe* = 17031.8, $p = 0.001$, as the match conditions were read faster than the mismatch conditions. However, there was no indication of an interaction between ambiguity and gender in any of the measures, *F*s < 1.

Reflexive Region: There was no significant main effect of ambiguity in any of the three measures, *F*s < 1. This strongly suggests that readers had completed reanalysis prior to fixating the reflexive. However, we did observe a significant GMME in a measure other than total time, as first pass time was longer in the mismatch conditions than in the match conditions, $F_1(1, 20) = 8.93$, *MSe* = 1026.0, $p = 0.007$; $F_2(1, 28) = 6.11$, *MSe* = 2257.5, $p = 0.02$. Note that this is the earliest gender effect in the current study, as all the gender effects in earlier regions occurred only in the total time measure. This is understandable because readers must process both the earlier matrix subject and this reflexive in order for a gender mismatch effect to arise. There was also a GMME in go-past time, $F_1(1, 20) = 12.25$, *MSe* = 27795.1, $p = 0.002$; $F_2(1, 28) = 23.66$, *MSe* = 16115.6, $p < 0.001$, and total time as well, $F_1(1, 20) = 49.57$, *MSe* = 7821.1, $p < 0.001$; $F_2(1, 28) = 49.42$, *MSe* = 10698.6, $p < 0.001$. Crucially, the interaction between ambiguity and gender was not significant in any of the three measures analysed (*F*s < 1). Moreover, simple effects analysis showed that the GMME was reliable in go-past and total time, both for the garden path conditions and also for the non-garden path conditions, all *p*s < .05. However, for first pass reading time, these simple effects were not fully reliable in either the garden path condition nor the non-garden path condition, probably due to lack of power (Garden Path: $F_1(1,20) = 1.79$, $MSe = 1095.6$, $p = .2$; $F_2(1,28) = 3.36$, $MSe = 2378.7$, $p = .08$, Non-garden Path: $F_1(1,20) = 5.79$, $MSe = 1432.8$, $p = .03$; $F_2(1,28) = 2.91$, $MSe = 2024.2$, $p = .1$).

End of Sentence Region: As with the prior region, there was no significant main effect of ambiguity in any of the dependent measures, *F*s < 1. Interestingly, in first pass time, there was a reverse GMME, $F_1(1, 20) = 12.17$, $MSe = 26053.9$, $p = 0.002$; $F_2(1, 28) = 12.65$, $MSe = 35005.9$, $p = 0.001$, as the mismatch conditions were actually read faster than the match conditions. In first pass time there was also a significant interaction between ambiguity and gender, in the analysis by participants, but it was only marginal in the analysis by items, $F_1(1, 20) = 4.72$, $MSe = 13685.1$, $p = 0.042$; $F_2(1, 28) = 3.76$, $MSe = 25968.5$, $p = 0.063$. However, the pattern of this interaction is not consistent with our previous predictions as it is driven by a larger *reverse* GMME in the ambiguous condition. In later measures of processing such as go-past time and total time, the GMME was in the standard direction with longer times in the mismatch condition: go-past time, $F_1(1,20) = 54.39$, $MSe = 211471.0$, $p < 0.001$; $F_2(1, 28) = 29.85$, $MSe = 501065.7$, $p < 0.001$, and total time, $F_1(1, 20) = 48.35$, $MSe = 27852.3$, $p < 0.001$; $F_2(1, 28) = 49.64$, $MSe = 33690.4$, $p < 0.001$. Additionally, for these later processing measures, the interaction between ambiguity and gender did not approach significance, *F*s < 1. Overall, the data pattern for this end-of-sentence region suggests that readers often made immediate regressions out of the region in the gender mismatch condition to check earlier material, leading to relatively short first-pass reading times and relatively long go-past times.

Extra Analysis (*Reflexive Plus Following Word*). Although the results at the reflexive region showed no evidence for an ambiguity-by-gender interaction, it is nevertheless possible that such an interaction might be found if the region were extended to include an extra word to the right of the reflexive[2]. Expanding the region in this way could increase power, as there will be more fixations on a two-word region than a single word region. In addition, the expanded region allows for the possibility that the critical interaction might be found slightly downstream of the critical word position. We therefore report analyses for the three eye-movement measures for a combined region consisting of the reflexive plus the immediately following word. The means for this combined region are given in Table 3 below:

Table 3

Statistical results for the expanded region were largely equivalent to those reported above for the reflexive region without the extra word. There was a main effect of gender matching in all measures, indicative of a gender mismatch cost: First pass: $F_1(1, 20) = 14.72$, $MSe = 7867.0$, $p < .01$; $F_2(1, 28) = 19.66$, $MSe = 7990.4$, $p < .001$; Go-past: $F_1(1, 20) = 26.02$, $MSe = 84472.9$, $p < .001$; $F_2(1, 28) = 70.21$, $MSe = 40587.1$, p = .001; Total Time: $F_1(1,20) = 86.02$, $MSe = 20871.4$, $p < .001$; $F_2(1,28) = 59.32$, $MSe = 40388.9$, $p < .001$). There was no main effect of Garden path in either first-pass or total time (all *p*s > .05). There was a marginal main effect of Garden path in go-past times, with longer reading times in the ambiguous condition, but this approached significance only by items ($F_1(1, 20) = 1.66$, $MSe = 72473.1$, $p = 0.213$; $F_2(1, 28) = 3.87$, $MSe = 43362.2$, $p = 0.059$). There was no sign of an interaction in either first-pass or total time (all *p*s > .3). However, the go-past times did show an interaction of Garden path by matching, but again, this was reliable by items only ($F_1(1, 20) = 1.58$, $MSe = 93668.4$, $p = 0.224$; $F_2(1, 28) = 6.28$, $MSe = 32870.7$, $p < .05$). Note, however, that the trend, if anything, was for a larger mismatch cost for the ambiguous than the unambiguous conditions (381 vs. 224 msec), which is the reverse of what would have been expected if the revision of the main

[2] We thank an anonymous reviewer for suggesting this analysis

clause structure had been incomplete at this point. Indeed, given that this interaction was not replicated in any other measure, and was not fully reliable by both F_1 and F_2, we are inclined to regard it as spurious.

Discussion

Two important findings emerged from this experiment. First, there were reliable effects related to the misanalysis and reanalysis of temporarily ambiguous structures, replicating many previous studies (Frazier & Rayner, 1982, 1987). Second, and more importantly, we found evidence that the parser successfully constructs a detailed hierarchical syntactic structure that supports reflexive binding, at least by the time the reflexive is encountered in the second verb phrase conjunct. The GMME in the reflexive region and the subsequent end-of-sentence region strongly suggest that the parser attempted to link the reflexive pronoun to its antecedent, the matrix subject NP. The slowdown observed in the mismatch conditions resulted from the parser's attempt to link the reflexive to the antecedent. As we have discussed, the online processing of reflexive binding is sensitive to the detailed hierarchical structure of the sentence. The fact that we observe the GMME strongly indicates that the parser constructs a detailed syntactic structure that supports the reflexive binding configuration (the structure of c-command or co-argumenthood). This, in turn, suggests that the antecedent NP was analyzed as the matrix Subject. If that NP had not been reanalyzed as the matrix subject, it would not have been structurally accessible to the reflexive pronoun, and thus there should not have been a GMME. Importantly, this penalty was seen in both garden path and non-garden path conditions, indicating that even in the garden path conditions, reanalysis processing eventually led to the antecedent NP being analyzed as the matrix subject prior to readers encountering the reflexive. Moreover, there was no main effect of ambiguity on the reflexive itself, indicating that the processing difficulty due to the garden path had largely subsided by this point in the sentence. Indeed, extra processing difficulty would have been expected in the ambiguous conditions if the globally correct analysis had not been available at this point, since without this analysis, the reflexive would lack an accessible antecedent. There were no signs of such an effect; therefore, the parser must have successfully reanalyzed the matrix subject NP, which had previously been misanalyzed as the embedded object.

The speed with which the syntactic reanalysis took place in the current study (within just a few words after disambiguation) suggests that the persistent effects of a garden-path sentence on answers to explicit comprehension probes (Christianson et al., 2001, 2006) are unlikely to be the result of incomplete syntactic parsing. This interpretation raises the possibility that misanalyses observed in previous work derive not from failed syntactic reanalysis, but from a failure to inhibit—or an inability to ignore—the erroneous initial syntactic structure. In other words, misanalyses appear to stem from the co-presence of two syntactic structures: the one built initially, and the one built during reanalysis. Experiment 2 was designed to test this possibility.

Experiment 2

As discussed in the Introduction, the traditional view of how garden-path ambiguities are processed does not predict any lingering effects of the misanalyzed structure on the processing of subsequent material, either because the initial parse was fully reanalyzed prior to the processor moving on to new material, or because the ambiguity was never detected at all. If, however, processing at either the syntactic or semantic level (or both) is incomplete prior to moving on to subsequent material, then one would expect lingering effects of the temporary ambiguity. The results of Experiment 1 suggest that syntactic reanalysis is carried out to an extent that is detailed enough to instantiate the c-command syntactic relation necessary for a reflexive pronoun to be coreferential with its antecedent. Note too that the c-command relation cannot be established by merely aligning phrases in a proper order; for c-command to be computed, the internal hierarchical structure of the phrases must be present and must be correct. However, it is possible that what does remain "good enough" in the processing of garden path sentences is the process of completely overwriting the initial, incorrect syntactic structure with the new, correct syntactic structure. In other words, reanalysis requires not only that the parser build the correct syntactic structure, but also that it get rid of the wrong one.

Experiment 2 was designed to examine reading patterns on text subsequent to the disambiguating region to determine the extent to which the initial interpretation is ignored or inhibited. Participants read pairs of sentences while their eye movements were monitored - for example, *While Frank dried off the truck that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog* (see (6) below for an illustration of an item in all the experimental conditions). The first sentence of each pair consisted of a subordinate-clause ambiguous syntactic structure that utilized either a reflexive absolute transitive (RAT) verb, e.g., *bathe, dress, dry off, scratch* (Christianson et al., 2001, 2006), or a reciprocal verb such as *hug*, *kiss*, and *meet* (Ferreira & McClure, 1997). RAT verbs are notable in that they are obligatorily semantically transitive, even when there is no overt object mentioned in the sentence. For example, *Anna dressed the baby* means that the baby was the syntactic object and patient of *dress*. But, *Anna dressed* means that she dressed herself (and is interchangeable with *Anna dressed herself*, with the overt reflexive pronoun), no matter how prominently the baby or even the baby's nakedness may be mentioned in the context. Reciprocal verbs share the property of absolute transitivity with RAT verbs in that they can appear with or without an overt object, but when they appear without an overt object, they can only be understood reciprocally, i.e., the direct object must be co-referential with the subject(s) (*The young couple hugged = The young couple hugged each other*). For this reason, if an overt object is omitted, then the subject must be syntactically (*the lovers*) or semantically (*the couple*) plural. The second sentence was identical in all conditions and was consistent only with the reflexive or reciprocal interpretation of the first sentence (Frank dries himself off, not the truck). The second sentence then allows us to probe the extent to which an incorrect initial parse lingers and interferes with the processing of new input.

We also introduced a plausibility manipulation much like the one used by Pickering and Traxler (1998) to explore how deeper semantic relationships influence lingering effects of garden path sentences (see (6) below). In this manipulation, the temporarily ambiguous noun phrase (NP) was either plausible or implausible as a patient object of the verb. For example, in (6) below, *the truck* is plausibly an object that can be dried, but *the grass* is not. The goal was to determine whether plausibility could block either the initial incorrect attachment of that NP as object of the subordinate verb – and thereby eliminate or dampen behavioral signals of syntactic reanalysis – or the lingering interpretation of it as patient of the verb. If the NP is implausible, then processing is predicted to be easier, because the parser will not be as tempted to interpret that NP as the object of the verb, and therefore the garden-path will be less severe or perhaps even nonexistent. Additionally, in the event that these implausible NPs are initially interpreted as the object of the verb, the implausibility of the resulting interpretation should make it easier to completely revise the structure during reanalysis.

In the first sentence, we expect to observe classic garden-path effects in reading times on the first sentence (cf. Frazier & Rayner, 1982), namely inflated reading times on the disambiguating region combined with increased regressions and re-reading times in the temporarily ambiguous compared to the unambiguous condition, as we saw in Experiment 1. In addition, consistent with Pickering and Traxler (1998), we predict that implausible NP objects will be less disruptive to reading overall, especially in garden path conditions.

In the second sentence, we expect that if full reanalysis (syntactic and semantic) takes place in garden-path sentences with RAT or reciprocal verbs, then during the reading of the second sentence we should observe no interference from the temporary and ultimately incorrect initial interpretations built during the initial processing of the preceding sentence. However, if reanalysis processes finish prior to completely over-writing the initial structure and inhibiting the original interpretation, we would expect to see longer reading times in the second sentence when referencing information that had been originally misparsed. Furthermore, if the expected interaction occurs whereby plausible garden paths yield an incorrect parse that is more resistant to attempts at reanalysis and or lead to a stronger memory trace than those created from implausible garden-path sentences, we should find stronger evidence for lingering competition between the initial parse and the revised parse in the critical region of the second sentence in the plausible garden-path condition. If, contrary to the results of Experiment 1, reanalysis were to result in the creation of an incomplete structural analysis of the first sentence, we would expect a global slowdown in reading of the second sentence—as it is difficult to process globally ungrammatical input (Munro & Derwing, 1995; Vasishth, Brüssow, Lewis, & Drenhaus, 2008)—rather than a specific slowdown at the critical region that references the misparse.

Finally, it is important to consider again the possibility that the parser does construct a detailed and accurate syntactic analysis during reanalysis, but only on a proportion of the trials. We considered such a possibility in Experiment 1 but failed to find a number of effects in our data that this alternative explanation clearly predicted (there was no main effect of ambiguity and no interaction between ambiguity and gender mismatch at the reflexive region). However, we may have failed to find such effects due to a potential lack of power. In Experiment 2 this alternative would predict that, on some proportion of trials, participants will have accurately completed reanalysis including completely expunging the incorrect initial structure and interpretation. For these trials we would expect no lingering difficulty in processing the next sentence (regardless of its content). However, on other trials reanalysis completely fails and the participants move on to interpret the content of the second sentence with the initial incorrect interpretation of the first sentence. For these trials we would expect that encountering the critical region of sentence 2 would initiate a second call for reanalysis of the first sentence given the semantic incompatibility of the second sentence with the garden path interpretation of the first sentence. If one assumes that reanalysis efforts are more likely to fail with plausible NPs than with implausible ones, this alternative explanation would make a similar prediction for reading times in the second sentence as the Good Enough-theoretic prediction we are making for the first sentence: greater processing difficulty in the critical region of the second sentence in the plausible than the implausible garden-path condition (relative to unambiguous controls). However, these two accounts make different predictions for the processing of the second sentence. If reanalysis completes prior to the total overwriting of initial syntactic and semantic representations, readers will arrive at the second sentence with multiple active representations: one decaying initial representation and one resulting from reanalysis. However, the alternative account predicts the presence of only one representation: either the initial incorrect interpretation or the correct interpretation that was created during reanalysis. Distinguishing between these two accounts may prove difficult. However, it seems reasonable to assume that readers who arrive at the second sentence with a single incorrect interpretation of the first sentence will need to reread aspects of the first sentence in order to arrive at a globally coherent interpretation of the two sentence phrase. Therefore, this account would likely predict an effect in go-past time for the critical region of the second sentence. In contrast, lingering misinterpretations of the first sentence can be resolved without the need to reread the first sentence and are likely to result in first-pass reading time effects.

Method

Participants. Twenty-eight adults from the University of California at San Diego community participated in the experiment. All had normal or corrected to normal vision, were native speakers of American English, and were naive to our research questions. They received extra credit for psychology classes or were paid eight dollars for their participation.

Apparatus. Eye-movements were recorded using the same equipment and in the same way as in Experiment 1.

Materials. The experiment had a 2 (garden path vs. non-garden path) by 2 (plausible vs. implausible embedded NP) within-participants design. The embedded NPs were matched for both length and frequency (Kučera & Francis, 1982). Stimuli consisted of 24 sentence pairs. Each pair could be seen in one of four conditions. Four lists of sentences were created using a Latin square design. An example stimulus in the four different conditions is shown in (6):

(6) Sample set Experiment 2 stimuli

a. *Non-Garden Path/Plausible*

While Frank dried off, the truck that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.

b. *Garden Path/Plausible*

While Frank dried off the truck that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.

c. *Non-Garden Path/Implausible*

While Frank dried off, the grass that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.

d. *Garden Path/Implausible*

While Frank dried off the grass that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.

Procedure. The procedure was identical to Experiment 1 except that comprehension questions were asked after a third of the items and there were 118 filler items interspersed, which were two-sentence experimental items from an unrelated experiment[3].

Results

Data exclusion criteria were the same as in Experiment 1: fixations shorter than 80 ms (1.98%), trials with fixations longer than 1200 ms (.003%), and data points that were three standard deviations above the condition mean (< 3%) were removed from the data record. For analysis,

[3] Fillers included some sentences with fronted adverbials that had licit direct objects in an attempt to counter-act any potential strategic effects.

sentences were divided into critical regions, as shown in Table 4. For each region, 2 (sentence structure) x 2 (plausibility) x 4 (list) (Pollatsek & Well, 1995) mixed design ANOVAs were performed, with participants (F_1) and items (F_2) as random effects. We analyzed the same three eye movement measure used in Experiment 1: first-pass time, go-past time (also called regression path duration; cumulated fixations before leaving the target region to the right), and total time. Significant results reported below were obtained at or below $p = .05$. The means of the various eye movement measures are given in Table 5. Though the most pertinent data are those of the critical region in sentence 2, we will present the reading data by order of the regions, as we did in our reporting of Experiment 1.

Table 4

Ambiguous Region: First pass times were shorter in the garden path condition (56ms); however, this difference was not significant $F_1(1,24) = 2.07$, $MSe = 42238.9$, p > .15; $F_2 < 1$. Consistent with Pickering and Traxler (1998), first pass time in the plausible noun phrase condition was 113ms shorter than in the implausible condition, $F_1(1,24) = 6.09$, $MSe = 50540.0$, $p < .05$; $F_2(1,20) = 4.59$, $MSe = 66283.8$, $p < .05$. In total reading time, participants spent 269 ms longer on the garden path structures $F_1(1,24) = 8.64$,$MSe = 234669.4$, $p < .01$; $F_2(1,20) = 8.79$,$MSe = 237067.0$, $p < .01$. There were no significant interactions, $ps > .05$.

Table 5

Disambiguating Region: In go-past time there was a significant effect of NP plausibility, with plausible patient NPs having longer times (238ms), $F_1(1,24) = 10.61$, $MSe = 149867.5$, $p < .01$; $F_2(1,20) = 7.74$,$MSe = 154125.2$, $p < .05$, and a significant effect of structure, with garden path structures having longer times (390ms) $F_1(1,24) = 18.98$, $MSe = 225173.9$, $p < 0.001$; $F_2(1,20) = 32.15$, $MSe = 127705.3$, $p < 0.001$. These main effects were qualified by a significant interaction between these factors in go-past time, as the effects of structure were larger when the NP had been a plausible patient of the verb, $F_1(1,24) = 14.87$, $MSe = 53532.2$, $p < 0.01$; $F_2(1,20) = 6.73$, $MSe = 121196.5$, $p < 0.05$. Total reading times on the disambiguating region were very similar to go-past times with a significant effect of NP plausibility (102ms), $F_1(1,24) = 6.94$, $MSe = 42045.5$, $p < 0.05$; $F_2(1,20) = 7.40$, $MSe = 29034.4$, $p < 0.05$, and a significant effect of structure (111ms), $F_1(1,24) = 8.30$, $MSe = 41264.5$, $p < 0.01$; $F_2(1,20) = 20.02$, $MSe = 19392.9$, $p < 0.001$. However, the interaction between these two factors was not fully significant in total reading time, $F_1(1,24) = 3.53$, $MSe = 25920.2$, $p > 0.05$; $F_2(1,20) = 4.62$, $MSe = 23266.5$, $p < 0.05$. No other effects were significant, *ps* > .05. The data pattern for both the go-past and total reading time measures are consistent with Pickering and Traxler (1998).

Opening Sentence 2 Region: Readers were slower in first pass time (56ms), go-past time (50ms), and total time (58ms) on the beginning of the second sentence when the NP in the first sentence had been a plausible patient of the verb but this effect was not fully significant in any of the measures, $F_1(1,24) = 3.36$, $MSe = 26210.9$, $p > 0.05$; $F_2(1,20) = 10.63$, $MSe = 7968.8$, $p < 0.01$, $F_1(1,24) = 2.12$, $MSe = 33472.3$, $p > 0.15$; $F_2(1,20) = 3.13$, $MSe = 12333.4$, $p > 0.05$, $F_1(1,24) = 3.07$, $MSe = 31398.9$, $p > 0.05$; $F_2(1,20) = 3.86$, $MSe = 14535.3$, $p > 0.05$, respectively. Notably, the garden path main effect was not significant in first pass, $F_1(1,24) = 1.40$, $MSe = 23094.7$, $p > 0.20$; $F_2 < 1$, go-past, $F_1 < 1$; $F_2(1,20) = 1.62$, $MSe = 11169.1$, $p > 0.20$, or total time, *F*s <1. Thus there is no evidence of a global slowdown in reading rate that would be indicative of a failure to create a complete syntactic structure for the first sentence. There was also no significant interaction between noun phrase plausibility and structure *ps* > .20.

Critical Region: This region of the second sentence included text that was consistent only with the reflexive interpretation of the first sentence. As such, increased fixation time would indicate difficulty with such an interpretation. For first pass time, neither the main effect of structure nor of noun phrase plausibility approached significance, *ps* > 0.15. As predicted, though, there was a significant interaction between structure and NP plausibility in first pass time, $F_1(1,24) = 4.63$, *MSe* = 3256.0, $p < 0.05$; $F_2(1,20) = 8.47$, *MSe* = 1880.7; $p < 0.01$. Participants spent more time initially reading this region if the first sentence had a garden-path structure, but only if the NP had been a plausible patient of the verb. Additionally, the simple effect of structure in the plausible condition was significant $F_1(1,24) = 5.26$, *MSe* = 4220.4, p < 0.05; $F_2(1,20) = 6.17$, *MSe* = 3859.5, $p < 0.05$; the simple effect of structure for the implausible conditions was not *Fs* < 1. There were no significant main effects or interactions for the go-past time and total time measures, *ps* > .05.

Final Region: There was a trend for longer reading times on the final region when the NP in the first sentence had been implausible as the patient of the verb, but this difference was not significance in any measure, *ps* > .05. No other effects approached significance for this region in any of the measures, *Fs* < 1.

Discussion

Experiment 2 investigated whether the structure associated with the initial syntactic misanalysis of a garden-path sentence lingered long enough to influence the processing of a follow-up sentence which assumed the correct, revised meaning. The data suggest that the semantic interpretation created from the initial syntactic structure persists and affects the processing of subsequent material, consistent with Sturt's (2007) notion of semantic persistence, as well as some versions of the Good Enough processing view (Christianson et al., 2001, 2006; Ferreira, et al., 2002; Ferreira & Patson, 2007). We furthermore argue that this lingering semantic interpretation requires the refining of the Good Enough Theory of sentence processing such that it be tied to the availability of semantic relationships within the initial misanalysis.

There were several noteworthy results from this Experiment 2. The first is the significant difference in first-pass reading times such that NPs that were implausible objects of the subordinate verb were read more slowly than plausible objects of the subordinate verb. This slowdown may be due to hesitancy on the part of the parser to integrate the implausible ambiguous NP into the current structure as the direct object of the subordinate verb. Next, a significant interaction was observed in go-past reading times on the disambiguating region of the first sentence. Assuming that go-past times are a measure of re-reading, and presumably reanalysis, the significant interaction observed between structure and plausibility indicates a greater reanalysis cost for NPs that are plausible patients of the subordinate verb. The explanation for this interaction can be found in both parallel (e.g., MacDonald et al., 1994) and serial (e.g., Frazier, 1987) parsing theories: plausible interpretations resist revision more than implausible ones.

The results of Experiment 2 discussed so far replicate those reported by Pickering and Traxler (1998), who also found earlier disruption with implausible objects and later disruption with plausible objects in the same sort of structure. In addition, the second sentence in the present experiment provides further insight into the relevant processes and sheds light on the extent of reanalysis. If the materials used here had consisted of only one sentence each, we would likely have concluded that all reanalysis processes had run to completion, and that readers had entirely expunged the initial, ultimately incorrect interpretation that, for example, "Frank dried off the truck." Instead, the significant interaction in the first-pass time on the critical region of the second sentence of our stimuli provides strong evidence for lingering interference from the initially built structure in which *the truck* was the patient of *dried off*. The results suggest that reanalysis of the first sentence did not fail: later measures on the critical sentence 2 region (e.g., go-past times) did not display effects of the manipulations in sentence 1, as would be expected if readers had failed to correctly reanalyse the first sentence (on all or a proportion of trials) and then tried again after receiving more information in the second. Nor was there a main effect of the first sentence's syntax alone on processing of the any of the regions of the second sentence (i.e. no global slowdown of reading rate associated with ambiguity). Instead, only in cases where a plausible NP was initially incorporated into the original, partial, ultimately incorrect structure for the first sentence was there a slowdown in first-past reading time for the second. It may be that when both syntax and semantics conspire to intensify the garden path, the original syntactic structure and semantic interpretation linger, but that the semantics ultimately cause the processing difficulty we observed in the second sentence. The data from our previous experiment indicate that the correct syntactic structure is built quickly during reanalysis, but those findings are compatible with a lingering of the initial, incorrect syntactic structure. Given that Sachs (1967) demonstrated the relative speedy decay of syntactic structure, and Christianson, Luke, and Ferreira (2010) reported semantic effects on structural priming, it seems reasonable to posit that semantic representations should linger longer than syntactic structures (see also Fodor, Bever, & Garrett, 1974). Nonetheless, it does appear that some aspects of the initial syntactic structure persist beyond the point at which that structure is still viable. It is also important to note that the critical region, which probed the interpretation of the ambiguous NP from the first sentence, was consistent with the globally correct analysis, and did NOT explicitly reactivate the incorrect parse. It has been suggested that the effect of lingering misinterpretations observed by Christianson et al. (2001) and Ferreira et al. (2001) is due to the reactivation of the interpretation by means of asking explicit comprehension questions (Tabor, Galantucci, & Richardson, 2004). The present result is inconsistent with this possibility, as there was no information that could have directly reactivated the misinterpretation. Therefore, in this study, we have strong evidence of lingering effects of initial misanalyses as they appear on-line and in the absence of any biasing information.

Our results require revisions to traditional models of sentence parsing. Serial, two-stage models of parsing (e.g., Frazier & Rayner, 1982) and parallel, one-stage models (e.g., MacDonald et al., 1994) generally assume that once a syntactic structure has been ruled out, it is quickly disabled so as not to continue to influence ongoing processing. However, the co-existence of the incorrect structure and the correct structure seems more compatible with some models of sentence comprehension that assume parallel processing - specifically, simultaneous consideration of more than one syntactic structure at some points during processing. We discuss these implications further in the next section.

General Discussion

The first experiment we reported showed that readers do perform complete syntactic reanalysis rapidly enough to elicit a Gender Mismatch Effect immediately following garden-path reanalysis. Experiment 2 showed that temporary ambiguity influenced on-line reading times on subsequent text that referenced the semantic interpretation of the temporarily ambiguous material. This second experiment is the first study to attempt to determine whether garden-path effects linger past the sentence containing the syntactic ambiguity, and the results are noteworthy because they reveal a significant semantic "hangover" from the original misparse of that ambiguity.

We view these results as being broadly consistent with the Good Enough theory. To date, the definition of Good Enough processing has been somewhat underspecified, but the present results point toward a more precise definition, as well as a parsing mechanism that predicts misinterpretation effects (Christianson, et al., 2001, 2006), structural perseveration effects (van Gompel et al., 2006), and semantic persistence (Sturt, 2007) of initial interpretations. In order for Good Enough processing to explain the results of Experiment 1, the theory must involve more than simple insensitivity of the parser to temporary syntactic ambiguity or the inability to perform complete syntactic reanalysis. As such, Good Enough processing does not seem to be characterized as an over-reliance on inference over structural information. Christianson et al. (2001, 2006) provided evidence against the over-reliance on inference, and the data from Experiment 2 are consistent with those findings: reanalysis was undertaken in the first sentence even if the ambiguous NP was plausible as an object of the first verb, but plausibility affected the extent to which this initial misanalysis was pruned. Based on the results from our two experiments, then, we believe that Good Enough processing refers to situations in which the parser carries on interpreting new input without having completely pruned interpretations that are no longer compatible with this input.

A possible mechanism for such lingering effects was proposed by Lau and Ferreira (2005; see also Ferreira, Lau, & Bailey, 2004), in an investigation of disfluency repair. They suggested that disfluent constituents are "overlaid" with repaired structure in a lexically guided tree-adjoining grammar (LTAG) parsing model (for details of this structural operation in an LTAG parsing model, see Ferreira et al. (2004)). Under this account, initial incorrect material is hypothesized to decay over time, and its continued presence competes with the correct, repaired structure, influencing subsequent interpretations and acceptability judgments. Lau and Ferreira (2005) speculated that the same sort of "overlay" procedure might hold for disfluency repair and garden path reanalysis alike. Instead of reanalysis ending before a legal and complete structure is created (Christianson et al., 2001), a revised structure may be overlaid on the existing, initial, ultimately incorrect tree. In the case of the sentences used particularly in the second experiment, when the matrix verb *was* is encountered, its simple syntactic tree is combined with that of the nearest subject NP (*the truck that was dark green*), and this combined structure is overlain on top of the tree associated with the existing misparse. Importantly, the structure and semantics of the NP that was initially misparsed as the object of *dried off* remain in the tree, and must itself be overlain with the structure containing the implicit reflexive *himself*. Figure 1 illustrates how this process leaves the ambiguous NP in its initial incorrectly parsed position. In LTAG, the initial misparse either decays over time or must be actively inhibited. In the case of decay, the initial misparse continues to attract attention from interpretive processes, competing with the implicit reflexive NP.

Figure 1

In Experiment 1, a gender mismatch cost was found at the reflexive region, regardless of whether misanalysis had initially occurred. In the LTAG account, this cost can be explained if we assume that the binding conditions between the reflexive (e.g. *himself*) and its antecedent (i.e. the main clause subject) apply to LTAG local trees. As the reflexive and the main clause subject are arguments of the same predicate, they would be considered to occur in the same local tree, regardless of whether this local tree has been incorporated into the representation via overlay (as in the ambiguous conditions) or via standard attachment operations (as in the unambiguous conditions).[4] Thus, a statistically equivalent mismatch cost is found when the reflexive mismatches in gender with its local subject, in both the ambiguous and unambiguous conditions.

The continued presence of the original misparse in this LTAG mechanism can predict the structural effects observed by van Gompel et al. (2006) and in turn the semantic persistence of the original interpretation (Sturt, 2007). The structural effects observed in the eye movement record of Experiment 2 above are attributed to continued, though weakening, activation of the original misparse (and its semantic interpretation) in which the ambiguous NP is still the direct object of the subordinate clause, perhaps via an episodic memory trace of one or the other or both (Kaschak & Glenberg, 2004). This activation exerts enough pressure on the processor, inducing lingering confusion as to what the ultimate interpretation is. Importantly, this confusion is not resolved before readers move onto the next sentence. If the subsequent sentence had not explicitly referred back to the previous one, the processing would indeed have been "good enough"; that is, the failure to remove the incorrect structure would not have been detected.

The theoretical approach advocated here is similar in spirit to self-organized parsing models, which have been proposed to account for local coherence effects associated with sentences such as (7) (Tabor & Hutchins, 2004;Tabor et al., 2004):

(7) *The coach smiled at the player tossed a frisbee by the opposing team*

---

[4] Examples involving Exceptional Case Marking verbs (e.g. *John believes himself to be clever*) are counter-examples to

Similarly to the LTAG account sketched above, Tabor & Hutchins (2004) and Tabor et al.'s Dynamic Self-Organized parsing (SOPARSE) model uses lexically anchored tree fragments that represent syntactic and semantic information, which are associated with open attachment sites specifying the possibilities for combination. In SOPARSE, processing difficulty can arise when there are multiple conflicting possibilities for combining a new word with the current syntactic representation. Processing difficulty associated with local coherence (such as that encountered at the word *tossed* in (7) above) is predicted to occur when combinatory possibilities resulting from purely local syntactic information conflict with those that result from top-down syntactic information. In (7), at a purely local level, the phrase *the player* provides a potential subject for the new input word *tossed,* an analysis that is also consistent with the subsequent phrase *the frisbee.* However, this attachment of *tossed* is inconsistent with its globally correct attachment as a past-participle, as part of the reduced relative clause modifying *the player.* The self-organizing architecture of SOPARSE allows these two mutually inconsistent attachments to be made simultaneously, leading to a temporary period of processing difficulty while the two attachments compete for activation. One might assume that SOPARSE can explain the results of our Experiment 2, as it allows the globally coherent and locally coherent analyses to co-exist, just like the LTAG account sketched above. Indeed, we believe that SOPARSE can account for the reading patterns in the first sentence of our Experiment 2, as these can be explained in terms of the competition between the locally coherent mis-analysis and the globally coherent correct analysis, around the point where the disambiguating information is processed---for example, more competition would be expected from the mis-analysis if it is supported by plausibility, predicting longer reading times in this condition, as we found. There is, however, a crucial difference between SOPARSE and the LTAG accounts. In SOPARSE, there is no mechanism to maintain the activation of the original misanalysis once the globally correct analysis has won the competition. In

---

the generalization that an anaphor has to occur in the same local tree as its antecedent. See Kallmeyer and Romero (2007) contrast, the LTAG account assumes an overlay function that allows the original mis-parse to remain in memory over an extended period of time, as it is continues to be present "underneath" the new and correct parse.

Because of this, we believe that it is less straightforward for SOPARSE to account for the results of the second sentence in Experiment 2. Here, in the reflexive region, we found competition effects relating to the interpretation of the reflexive, which probed the globally correct analysis of the first sentence. It would be difficult to account for this effect in terms of sustained competition from the first sentence, as the results from the first region of the second sentence indicated that any initial competition related to the garden path had apparently already subsided. Moreover, as mentioned above, processing difficulty in SOPARSE arises when the current input word has multiple competing attachments (one of which may be merely locally coherent). However, the reflexive in the second sentence of our Experiment 2 has only one possible attachment, and the pattern of processing difficulty can only be explained in terms of the interpretation that is yielded by this attachment, and its relation to the first sentence.

An alternative to the Good Enough/LTAG account of our results appeals to activity in the memory system, which takes place as sentences are processed, analyzed, and reanalyzed. Consider once again the sentence *While Anna dressed the baby played in the crib*. Previously, we suggested that when a syntactic parse is built incorporating *the baby* as the grammatical object of *dressed*, a proposition is incrementally established representing the idea that *Anna*, the agent, dressed *the baby*, the patient. This proposition exists in working memory, and is linked to the parse from which it was derived. When the parser encounters *played* and initiates syntactic reanalysis, the direct object *the baby* is detached from the subordinate clause and made the subject of the main clause. Again, incrementally, the semantic processor creates the interpretation that the baby is playing in the crib. In addition, the semantic processor should now revise the proposition built earlier, which represents the

for an LTAG approach to anaphor binding that covers such cases.

baby as a patient. The argument *the baby* must be deleted from the proposition and some type of reflexive interpretation must be built instead. Assuming that this semantic step of revising the initially built proposition is undertaken, the comprehension system is now faced with a problem: It owns two different sets of propositions for the same sequence of words. For semantic reanalysis to be successful, the comprehension system must maintain the bindings between syntactic parses and propositional representations so that it knows to edit the representations created during the initial misparse, and must also make sure to select the interpretations corresponding to the correct parse, not the initial parse. If we assume that memory is fallible (Schacter, 1999), then it seems reasonable to expect that binding errors will sometimes occur, resulting in what we have termed Good Enough interpretations. On this view, the working memory effects observed in previous work (Christianson et al., 2006) are expected as well, since individuals with greater working memory capacities are likely to maintain multiple interpretations more successfully, as well as the accurate bindings between those interpretations and the structures from which they arose.

A recent approach to sentence processing, described by Levy (2008) and Levy et al. (2009) also predicts downstream effects of earlier uncertainty related to the input string. Levy et al. (2009) implement this prediction by combining top-down structural expectations with bottom-up (i.e., incremental) input filtered through noise, resulting in a Bayesian inference about more or less uncertainty regarding the veracity of the input. In other words, readers may not always be certain that what they think they have read is really what was actually written. New input can cause readers to re-evaluate their beliefs about prior context. Additional processing is required in order to make a large shift in the probability distributions of the potential parses for the previous input. While the analyses in this model are always grammatically faithful to the supposed input string, the beliefs about the prior context can include analyses that are not consistent with the veridical perceptual input. For example, the real input string might be "Word1 Word2 Word3", but there might be some non-veridical neighbors of this input string (e.g. "Word1, WordX, Word2, Word3") whose syntactic analyses receive some probability mass in the distribution. We shall refer to an analysis as "veridical" if it corresponds to the actual word string, and "non-veridical" if it doesn't. The model then predicts processing difficulty when an incoming word provides evidence for either a veridical or a non-veridical analysis that had previously had a relatively low probability. In our Experiment 2, the ambiguous conditions have the following veridical parses and non-veridical counterparts:

Plausible (veridical)

[While Frank dried off][the truck that was dark green was peed on by a stray dog.]

Plausible (non-veridical)

[While Frank dried off the truck that was dark green] [**it** was peed on by a stray dog.]

Implausible (veridical)

[While Frank dried off][the grass that was dark green was peed on by a stray dog.]

Implausible (non-veridical)

[While Frank dried off the grass that was dark green] [***it*** was peed on by a stray dog.]

Assuming that semantics contributes to the probability of selecting an analysis, the non-veridical analyses would have a higher probability in the plausible than the implausible conditions. Then when readers reach the critical region of the second sentence, they obtain additional evidence favoring the veridical analysis over the non-veridical one, which in turn would result in a larger shift of probabilities, and therefore processing cost, in the plausible than the implausible garden path condition. Therefore, in principle, the Levy et al. model may be capable of predicting the lingering interactive effects we obtained in experiment 1. However, this model was only specified for a single sentence and it is unclear how the model could be extended to handle effects of uncertainty in subsequent sentences. For instance, the current model does not include any decay function for uncertainty. It seems unlikely that the lingering effects we saw in Experiment 2 would remain indefinitely. In fact, an interesting question for future research is how long this proposed uncertainty lasts.

Further research will be required to decide between the explanations of our data that we have offered here, the one appealing to LTAG and the notion that initial syntactic structures decay over time and create interference until they are gone, and the other appealing to the fallibility of memory processes. Of course, it is certainly possible that both accounts are correct: Syntactic structures (or their associated interpretations) might not immediately be erased, and therefore they may continue to exert some effect until they decay entirely; additionally, memory for interpretations built at different stages and associated with different structures is likely to be fallible. For now, we believe the contribution of the present work is to demonstrate that lingering semantic interpretations do not reflect the parser's failure to revise the syntactic structure properly and to create a fully articulated grammatical form. Therefore, it appears that syntactic reanalysis is fairly complete and leads to the creation of a structure that is faithful to the input. In other words, it is not the representations that are just "good enough"; it is the processes that create the representations and fail to inhibit those representations that are no longer faithful to all the available lexical information.

References

Badecker, W. & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 748-769.

Chomsky, N. (1981). *Lectures on government and binding: The Pisa Lectures*. Holland: Foris Publications. Reprint. 7th Edition. Berlin and New York: Mounton de Gruyter, 1993.

Christianson, K., Hollingworth, A., Halliwell, J.F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368-407.

Christianson, K., Luke, S.G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36,* 538-544.

Christianson, K.,Williams, C.C., Zacks, R.T., & Ferreira, F. (2006). Misinterpretations of garden-path sentences by older and younger adults. *Discourse Processes, 42,* 205-238.

Crain, S. & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D.R. Dowty, L. Karttunen, and A.M. Zwicky (Eds.), Natural language parsing: Psychological, computational, and theoretical perspectives (pp. 320-358). Cambridge, UK: Cambridge UP.

Drieghe, D., Rayner, K., & Pollatsek, P. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *The Quarterly Journal of Experimental Psychology,* 61(8), 1239-1249.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11-15.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30,* 3-20.

Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing. Journal of *Memory and*

Language, 25, 348-368.

Ferreira, F. & Henderson, J.M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30*, 725-745.

Ferreira, F., Lau, E. F., & Bailey, K.G. D. (2004). Disfluencies, language comprehension,and tree adjoining grammars. *Cognitive Science*, *28*, 721-749.

Ferreira, F. & McClure, K. K. (1997). Parsing of garden-path sentences with reciprocal verbs. *Language and Cognitive Processes, 12*, 273-306.

Ferreira, F., & Patson, N. (2007). The good enough approach to language comprehension. Language and Linguistics Compass, 1, pp 71-83.

Fodor, J. D. and Inoue, A. (1998). Attach anyway. In J. D. Fodor and F. Ferreira (Eds.), Reanalysis in Sentence Processing, 101-141. Dordrecht: Kluwer Academic Publishers.

Fodor, J., Bever, A., Garret, T.G. (1974). *The psychology of language: An introduction to Psycholinguistics and generative grammar*. New York: Mcgraw-Hill.

Francis, W. N. & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*, pp. 559-586. Hillsdale, NJ: Erlbaum.

Frazier, L., and Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.

Frazier, L. & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language, 29*, 181-200.

Gass, S., & Varonis, E. (1984). The effect of familiarity on comprehension of the non-native speech. *Language and Speech, 34*, 65-89.

Gibson, E. (1991). (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. *Unpublished Ph.D. Thesis.*

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, and W. O'Neil (Eds.), *Image language, brain: Papers from the First Mind Articulation Project Symposium*, 95-126. Cambridge, MA: MIT Press.

Kallmeyer, L. & Romero, M. (2007). Reflexives and Reciprocals in LTAG. In Geertzen, J., Thijsse, E., Bunt, H., and Schiffrin, A. (Eds) *Proceedings of the Seventh International Workshop on Computational Semantics (IWCS-7), 271-282,* Tilburg

Kaschak, M. P. & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, *133*, 450-467.

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58, 239-261.

Lau, E. F. & Ferreira, F. (2005). Lingering effects of disfluent material on comphrehension of garden path sentences. *Language and Cognitive Processes*, *20*, 633-666.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. Cognitive Science, 29: 375-419.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126-1177.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences, 106*, 21086-21090.

MacDonald, M.C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157-201.

MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676-703.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 45*, 73-97.

Pickering, M.J. & Traxler, M.J. (1998) Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 940-961.

Pollatsek, A. & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 785-794.

Reinhart, T. & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, 24, 657-720.

Runner, J.T., Sussman, R.S., Tanenhaus, M.K. (2006). Processing reflexives and pronouns in picture noun phrases. *Cognitive Science: A Multidisciplinary Journal, 30,* 193-241.

Runner, J.T., Sussman, R.S., Tanenhaus, M.K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition, 89,* B1-B13.

Sachs, J.S. (1967) Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437-442.

Schacter, D. (1999) The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist, vol 54, 3*, 182-203.

Staub, A. (2007a). The return of the repressed: Abandoned parses facilitate syntactic reanalysis. *Journal of Memory and Language*, 52, 299-323.

Staub, A. (2007b). The parser doesn’t ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 550-569.

Sturt, P. 2003. 'The Time-Course of the Application of Binding Constraints in Reference Resolution', Journal of Memory and Language, 48 (3): 542-62.

Sturt, P. (2007). Semantic re-interpretation and garden-path recovery. *Cognition, 105*, 477-488.

Swets, B., Desmet, T., Clifton, C., Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, 36, 201-216.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355-370.

van Gompel, R.P.G., Pickering, M.J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory & Language*, *55*, 335-36.

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

Acknowledgments

We would like to thank Danika Paskvan for her excellent proofreading of this manuscript.

This research was supported by an NFS grant (BCS-0847533) to Kiel Christianson. Correspondence should be sent to Timothy J. Slattery (slattery@southalabama.edu).

FIGURE CAPTION

FIGURE 1: The syntactic parse tree at the point after which the disambiguating region has been encountered and its structure overlain on the existing sentence structure. The gray portion of the tree represents the structure of the ambiguous NP after overlay, as it begins to decay over time. The bolded portion represents the structure that is overlaid onto the existing tree upon reaction the disambiguation and reanalyzing accordingly.

TABLE 1: Critical regions of test sentences in Experiment 1

Region 1	Region 2
Connective plus subordinate subject	**Subordinate verb plus matrix subject**
After the bank manager	*telephoned David's father*

Region 3	Region 4	Region 5	Region 6
Disambiguation	**Spill-Over**	**Reflexive**	**Final Region**
grew	worried and gave	himself	approximately five days to reply.

Means for all measures and regions are reported in Table 2.

TABLE 2: *Reading times in msec. (SE), Experiment 1*

Region	Garden Path Match	Non-GPath Match	Garden Path Mismatch	Non GPath Mismatch
Region 1: *After the bank manager*				
First Pass	828 (54)	819 (53)	864 (50)	791 (52)
Total Time	1701 (116)	1392 (91)	1687 (139)	1463 (90)
Region 2: *telephoned David's father*				
First Pass	481 (31)	468 (26)	508 (28)	483 (32)
Go-Past	735 (50)	612 (36)	737 (52)	641 (47)
Total Time	1139 (98)	806 (50)	1288 (105)	1044 (76)
Disambiguating: *grew*				
First Pass	316 (13)	293 (14)	322 (17)	282 (12)
Go-Past	592 (50)	350 (17)	563 (55)	369 (21)
Total Time	534 (31)	433 (23)	584 (34)	495 (28)
Spill-over: *worried and gave*				
First Pass	375 (20)	338 (14)	368 (18)	346 (21)
Go-Past	539 (38)	422 (21)	599 (34)	417 (23)
Total Time	558 (34)	476 (23)	640 (37)	569 (32)
Reflexive: *himself*				
First Pass	250 (10)	237 (11)	262 (9)	263 (9)
Go-Past	314 (30)	282 (16)	434 (50)	400 (28)
Total Time	338 (22)	326 (17)	454 (29)	464 (30)
End of Sentence*: approximately five days to reply.*				
First Pass	901 (60)	832 (59)	735 (41)	769 (60)

Total Time	1167 (63)	1239 (69)	1425 (84)	1456 (85)

Note: All times are in milliseconds. Standard errors appear in parenthesis. Go-past times are not presented for Region 1 (these are equal to first pass times) or the End of Sentence region (there is no text past this region).

Table 3: Means and standard errors for the combined region (reflexive plus immediately following word)

	Garden Path Match	Non-GPath Match	Garden Path Mismatch	Non Gpath Mismatch
First Pass	438 (21)	449 (22)	520 (27)	537 (25)
Go-past	551 (33)	559 (36)	932 (103)	782 (57)
Total Time	624 (35)	656 (33)	892 (55)	934 (50)

TABLE 4: Critical regions of test sentences in Experiment 2

Ambiguous Region	**| Disambiguation**
While Frank dried off(,)the truck (grass) that was dark green	**|***was peed on by a stray dog.*

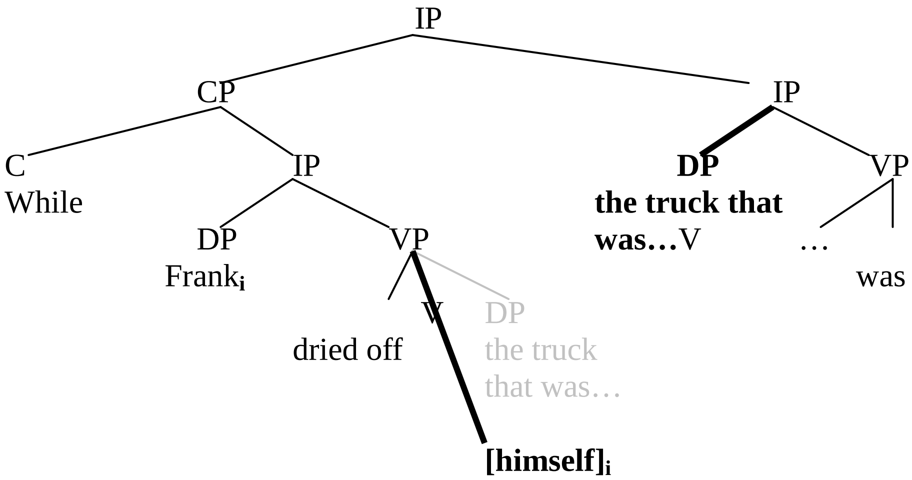
Opening S2 Region	**| Critical Region**	**| Final Region**
Frank quickly finished drying	**|** *himself off*	| *then yelled out the window at the dog.*

TABLE 5. *Reading times in msec. (SE), Experiment 2*

Region	Garden path Plausible	Non-garden path Plausible	Garden path Implausible	Non-garden path Implausible
Ambiguous: *While Frank dried off(,)the truck (grass) that was dark green*				
First pass	1811 (95)	1913 (100)	1962 (105)	1971 (118)
Total time	2758 (174)	2434 (149)	2657 (177)	2443 (160)
Disambiguating: *was peed on by a stray dog.*				
First pass	657 (39)	680 (48)	718 (44)	685 (48)
Go-past	1668 (163)	1108 (87)	1261 (108)	1039 (82)
Total time	1094 (87)	927 (63)	935 (64)	882 (63)
Opening S2: *Frank quickly finished drying*				
First pass	864 (50)	854 (40)	832 (45)	774 (41)
Go-past	910 (52)	894 (53)	859 (42)	845 (44)
Total time	964 (52)	980 (55)	925 (43)	902 (40)
Critical: *himself off*				
First pass	341 (20)	301 (15)	310 (18)	316 (13)
Go-past	381 (26)	360 (24)	352 (21)	342 (18)
Total time	418 (23)	371 (20)	393 (19)	394 (20)
Final: *then yelled out the window at the dog.*				
First pass	895 (64)	896 (70)	938 (68)	953 (70)
Total time	1028 (60)	1026 (73)	1078 (73)	1074 (67)

Note: All times are in milliseconds. Standard errors appear in parenthesis. Go-past times are not presented for the ambiguous region (these are equal to first pass times) or the final region (there is no text past this region).

FIGURE 1


IP
CP
IP
C
While
IP
DP
Frank$_i$
VP
V
dried off
DP
the truck
that was…
[himself]$_i$
DP
the truck that
was…
V
VP
…
was

APPENDIX A

Items from Experiment 1

1. Because the party guests applauded|,| Samuel's |son/daughter| danced clumsily and exposed himself unknowingly to ridicule.

2. When the teacher taught|,| Martin's |grandson/granddaughter| shrieked and hid himself nervously under the desk.

3. While the instructor was lecturing|,| Martin's |nephew/niece| shifted and bumped himself hard against the desk.

4. While the painter was sketching|,| Matthew's |brother/sister| fainted and lowered himself unsteadily onto the chair.

5. After the robber left|,| Melvin's |son/daughter| crouched and untied himself methodically from the table leg.

6.Although the dealer cheated|,| Mark's |grandfather/grandmother| smiled knowingly and rearranged himself carefully in the uncomfortable chair.

7. Because the model hurried|,| Douglas's |brother/sister| became irritated and ordered himself champagne and a dish of caviar.

8. Even though the professor lectured|,| David's |son/daughter| became bored and told himself always to check the instructor ratings before taking a class.

9. Because the teenager was starving|,| Andrew's |nephew/niece| panicked and gave himself permission to contact the authorities.

10. After the security guards stopped|,| Richard's |grandfather/grandmother| complained and argued himself cleverly out of the traffic ticket.

11. After the social workers visited|,| Oliver's |uncle/aunt| swore and drank himself slowly into a stupor.

12. After the bank manager telephoned|,| David's |father/mother| grew worried and gave himself approximately five days to reply.

13. After the midges attacked|,| Eddie's |son/daughter| screamed loudly and tore himself frantically out of the sleeping bag.

14. While the tennis player raced|,| Fred's |nephew/niece| cheered from the sidelines and made himself completely hoarse from all the shouting.

15. After the baby calmed down|,| Gerald's |grandfather/grandmother| sighed and poured himself numerous cups of strong tea.

16. While the salesclerk undressed|,| Harold's |brother/sister| winked and told himself immediately to visit that shop more often.

17. While the police were searching|,| Bill's |sister/brother| trembled and covered herself tightly with a thick blanket.

18. While authorities were investigating|,| David's |daughter/son| cried and cursed herself terribly for committing the crime.

19. As agents approached|,| James's |sister/brother| dashed and hid herself quickly in

the crowd of pedestrians.

20. As the parents bathed|,| Daniel's |daughter/son| slept and gave herself quite a
good rest.

21. When the gymnasts trained|,| Greg's |niece/nephew| sighed and dressed herself reluctantly in
a sweatshirt.

22. When the team was training|,| Benjamin's |niece/nephew| yawned and amused herself thoroughly by
thinking about dinner.

23. Before the nurse checked|,| William's |aunt/uncle| shivered and warmed herself happily with
a down jacket.

24. Before the friends left|,| Dean's |aunt/uncle| arrived and introduced herself politely to
all of the guests.

25. While the children watched|,| Jane's |sister/brother| wept and turned herself quickly into
an object of ridicule.

26. While the journalists photographed|,| Jeffrey's great |aunt/uncle| dozed and confused herself severely when
he later woke up.

27. While the historians interviewed|,| Helen's great |grandmother/grandfather| listened and enjoyed herself tremendously
with memories of old times.

28. After the customers paid|,| Peter's |aunt/uncle| smiled and considered herself extremely skillful
in the business world.

29. Before the committee met|,| Colin's |daughter/son| frowned and prepared herself fairly reluctantly for
the ordeal.

30. Whenever the children woke|,| Maria's |sister/brother| sighed and dragged herself slowly out
of bed to deal with it.

31. Before the teachers instructed|,| Tom's |granddaughter/grandson| fainted and declared herself clearly to
be unfit for learning.

32. After the school phoned|,| Jimmy's |mum/dad| frowned and poured herself several measures
of neat gin.

Appendix B

Items from Experiment 2

1. While Frank dried off|,| the |truck/grass| that was dark green was peed on by a stray dog.
Frank quickly finished drying himself off then yelled out the window at the dog.

2. As Betty woke up|,| the |neighbour/airplane| from Atlanta sat motionless.
It was not easy for Betty to wake herself up so she had two alarm clocks.

3. While the thief hid|,| the |jewel/guard| from the gallery could be seen on the security camera.
The thief hid himself in a place where the cameras couldn’t see him.

4. While the baboon groomed|,| the |chimp/tiger| at the Zoo was being photographed.
The baboon finished grooming himself then climbed up a tree and out of sight.

5. While Anna dressed|,| the |girl/home| that was stylish appeared on TV.
It seemed Anna always dressed herself while watching television.

6. While the boy washed|,| the |dog/sun| that was hot hid behind the trees.
The boy always washed himself before eating lunch.

7. While the jockey settled down|,| the |filly/wager| that was a long shot caught the gamblers' attention. The jockey always had to settle himself down before a big race.

8. While the mother undressed|,| the |child/phone| that was annoying made a racket.
It seemed the mother could never undress herself in peace.

9. While Jim bathed|,| the |baby/chair| that was wobbly fell with a thud.
The sound startled Jim and he got out of the bath to see what had happened.

10. While the nurse shaved|,| the |patient/surgeon| who was exhausted tried to nap.
The nurse shaved her legs before leaving the hospital because she had a date that night.

11. While the girl scratched|,| the |cat/bus| that was gray turned around slowly.
The girl scratched herself a lot since she walked through poison ivy.

12. While Janet calmed down|,| the |class/radio| that was usually noisy suddenly went silent.
Janet needed to calm herself down or she would lose control.

13. While Dan and Tim fought|,| the |bully/cloud| that was threatening loomed over them.
Tim and Dan fought each other often despite being brothers.

14. As Jane and Mary met|,| the |men/car| from Florida drove past them.
Whenever Mary and Jane met each other they couldn’t stop talking.

15. As the Finn and the Cuban raced|,| the |Italian/referee| who was overweight started to sweat.
The Cuban and Finn raced each other at nearly every track meet.

16. While Bill and Sue hugged|,| the |boy/art| from Moscow arrived at the museum.
Sue and Bill hugged each other more often than most married folks.

17. While Jill and Joe cuddled|,| the |kitten/stereo| that was small played quietly.
Joe and Jill liked to cuddle each other every night.

18. As Ed and Bea kissed|,| the |baby/snow| that was heavy fell gently to the ground.
When Bea and Ed kissed each other they forgot everything else.

19. As the duke and knight battled|,| the |prince/infant| who was dressed in red watched in awe. The knight and duke would battle each other to the death if no one intervened.

20. While Jodi and Liz embraced|,| the |girl/bird| with the pretty eyes drank some water. Liz and Jodi embraced each other after many years apart.

21. As the teacher and lawyer debated|,| the |politician/courtroom| that was tense paid close attention. The lawyer and teacher debated each other on the controversial issue.

22. While the doctor and dentist dated|,| the |nurse/puppy| with the cute nose acted shy. The dentist and the doctor dated each other for a year before everyone got used to it.

23. As the guard and officer wrestled|,| the |thief/stool| with the broken leg toppled over and rolled down the stairs. The officer and guard wrestled each other into a state of exhaustion.

24. As Mark and Janice touched|,| the |lamp/moon| that was pink glowed brightly. Mark and Janice swore that when they touched each other the earth moved.