Effects of Merely Local Syntactic Coherence on Sentence Processing

Whitney Tabor and Bruno Galantucci
University of Connecticut

Daniel Richardson
Cornell University

In press, *Journal of Memory and Language*

Please address correspondence to: Whitney Tabor
Department of Psychology, U-20
University of Connecticut
Storrs, CT 06269

860 486 4910 (phone)
860 486 2760 (fax)

tabor@uconnvm.uconn.edu

Abstract

A central question for psycholinguistics concerns the role of grammatical constraints in online sentence processing. Many current theories maintain that the language processing mechanism constructs a parse or parses that are grammaticality consistent with the whole of the perceived input when it processes each word. Several bottom-up, dynamical models make a contrasting prediction: partial parses which are syntactically compatible with only a proper subpart of the input are sometimes constructed, at least temporarily. Three self-paced reading experiments probed for interference from such locally coherent structures. The first tested for a distracting effect of irrelevant Subject-Predicate interpretations of Noun Phrase-Verb Phrase sequences (*The coach smiled at the player tossed a frisbee*) on reading times. The second addressed the question of whether the interference effects can be treated as lexical interference, instead of involving the formation of locally coherent syntactic structures. The third replicated the reading time effects of the first two experiments with grammaticality judgments. We evaluate the dynamical account, comparing it to several plausible approaches that also predict effects of local coherence, and arguing against several accounts which rule out the formation of merely locally coherent structures.

Keywords: parsing, sentence processing, garden-path, syntactic ambiguity, dynamical system, self-organization, bottom-up processing

INTRODUCTION

Much has been learned about the nature of sentence processing by studying temporarily ambiguous sentences like (1) – (3):

(1) The mechanic maintained the truck was working beautifully.

(2) The cop arrested by the detective was chagrined.

(3) The cook stirred the soup with the tomatoes.

Most current theories of sentence processing interpret these results by assuming that the processor builds sentence analyses incrementally, fitting each successive word into the current partial-parse. Difficulty is expected to arise when several parses are sanctioned by the grammar and the reader either (a) initially makes a wrong choice (serial models) or (b) activates multiple parses in parallel with the wrong one ranked highest and must revise the ranking (parallel models). Theories of ambiguous parsing in this class include the standard Garden Path Theory (Ferreira & Clifton, 1986; Ferreira & Henderson, 1990; Frazier & Rayner, 1982; Frazier, 1987; Frazier & Clifton, 1996), Reanalysis as a Last Resort theories and monotonic parsing theories (Fodor & Frazier, 1980; Frazier & Clifton, 1998; Gorrell, 1995; Phillips, 1998; Schneider & Phillips, 2001), Race-Based Parsing theories (Traxler, Pickering, & C. Clifton, 1998; vanGompel, Pickering, & Traxler, 2000, 2001), Syntactic Prediction Locality Theory (Gibson, 1998; Grodner, Gibson, & Tunstall, 2002), parallel parsers which use the input to assign weights to various grammatical options (Gorrell, 1987; Gibson, 1991; Jurafsky, 1996b; Hale, 2002).

Here, we consider an alternative possibility: it may be that local syntactic coherence in the input can result in the construction of syntactic analyses which are inconsistent with the global syntactic context. This paper asks if such errant local parses can have a detectable influence on the time course of processing.

By "local syntactic coherence in the input" we mean sequences of two or more words in the text stream which form a phrase or part of a phrase that cannot be grammatically unified with the parse of the preceding words in the sentence. An example is provided in (4).

(4) The coach smiled at the player tossed a frisbee by the opposing team.

In (4), "the player tossed a frisbee" has a locally coherent interpretation (both syntactically and pragmatically) as an active clause. But, in the context of the sentence, which strongly favors interpretation of "the player" as the direct object of "smiled at", "tossed a frisbee" can only be grammatically parsed as a reduced relative clause modifier of "player". We wondered if readers would nevertheless be distracted by the active interpretation. In particular, we

Many thanks to Ino Flores d'Arcais, Katherine Finerty, Carol Fowler, Edward Gibson, Sean Hutchins, Leonard Katz, Neal Pearlmutter, Aaron Schultz, Michael Tanenhaus, and Michael Spivey for helpful contributions and/or comments. Particular thanks to Garvin Boudle, Edward D'Agata, and David Perkowski for computer support. This research was supported by NIH Grant 5 T32 MH19389 and by University of Connecticut Research Foundation Grant FRS 444078.

investigated the hypothesis that readers would read more slowly at the ambiguous verb "tossed" and following words in (4) compared to a control case like (5) where there is less distracting local coherence.

(5) The coach smiled at the player who was tossed a frisbee by the opposing team.

If such a contrast obtains, we are aware of two broad classes of explanations: Local Coherence Accounts and Self-Consistent Parse Accounts.

Local Coherence Accounts.

In Local Coherence Accounts (LCAs), the contrast between the two examples comes about because the parser parses "the player tossed the frisbee" in (4) as an active clause, but finds no such parse in (5).

LCA1. One kind of Local Coherence Account stems from an approach to parsing that we refer to as "Self-Organized" parsing (Kempen & Vosse, 1989; MacDonald, Pearlmutter, & Seidenberg, 1994; Stevenson, 1994; Stevenson & Merlo, 1997; Stevenson, 1997, 1998; Vosse & Kempen, 2000; Tabor & Hutchins, in press). In Self-organized Parsing, words introduce tree-fragments with open attachment sites. These seek to combine with one another in ways that satisfy the constraints specified by their attachment sites. The strengths of the bonds which hold the parse together are continuous-valued and they wax and wane as a function of how much the input supports them. Under some versions of self-organization (e.g., Tabor & Hutchins, in press), if a group of later-arriving words can form a felicitous syntactic combination that is inconsistent with the current global parse, then this combination will form, at least weakly, and compete with the main parse. LCA1 predicts that (4) will induce strong competition from such a local formation at and immediately following the second verb, thus slowing reading, while examples like (5) will very little competition of this form, and reading will proceed more rapidly.

LCA2. Another kind of Local Coherence Account assumes that a module of the sentence processor operates on a buffer that contains several words. The output of this module is passed to the parser proper. If the preprocessing were done on the words "the player tossed the frisbee" in (4), then the module might output an active parse that would conflict with the main parse, slowing reading. Frazier and Fodor (1978) is an early proposal to transmit information to the parser via a fixed-width buffer. A more current and implemented proposal using n-grams is described in Corley and Crocker (2000), Crocker and Brants (2000).

LCA3. A third kind of Local Coherence Account assumes that the parser fails to find a coherent global parse on its first pass through sentences like (4) and (5). It then attempts to salvage at least a partial parse from the input. If it succeeds in building a parse fragment, as it does in (4) but not in (5), then higher reading times result, presumably because of the extra mental load imposed by this second parse-construction process. There are a number of reasons why the parse failure posited by LCA3 might occur. One plausible one is that the intended parses of (4) and (5) involve the extraction of a Recipient argument from a clause headed by a ditransitive verb. This kind of construction may be ungrammatical for some speakers of English.

Self-Consistent Parsing Accounts.

Many current theories of sentence processing assume that parsing is guided by a mechanism which has access to a full grammar of the language and that the grammar is used to constrain the interpretation of each incoming piece of information. While such accounts cannot attribute effects to mere local coherence, there are a number of reasons why readers with such grammatically self-consistent parsers might find the intended parse harder to construct in (4) than in (5). We identify one of these here, which we call Self-Consistent Parsing Account 1 (SCPA1), and several more below, the latter being motivated by some of the experimental findings we describe.

SCPA1. It might be the case that any contrast that exists between (4) and (5) is due entirely to the difficulties associated with processing reduced relative clauses. For example, the presence of grammatical cues in the form of the function words "who" and "was" may make it easier for the reader to build a correct parse in (4) than in (5), where there are fewer grammatical cues. This guiding effect of the function words might have a statistical basis (e.g., Corley & Crocker, 2000; Crocker & Brants, 2000; Hale, 2002; Jurafsky, 1996b; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Tabor, Juliano, & Tanenhaus, 1997)—the probability of a verb occurring as a past participle may be higher after "who was" than after a bare noun. Alternatively, in the absence of the function words, the parser might fail to construct relevant abstract phrase structure nodes until it is forced to by the arrival of the passive participle, resulting in a more complex tree-building operation at the participle in (4) than in (5) (cf. Frazier, 1978).

Overview

We conducted three experiments to examine the question of whether merely locally coherent material gets parsed (as per the LCA accounts and contra the SCPA accounts).

EXPERIMENTS

Experiment 1

Experiment 1 used Self-Paced Word-by-Word reading (see description below) to find out if an effect of Local Coherence could be distinguished from difficulty independently associated with processing reduced relative clauses.

Method

Participants.

Forty-seven subjects were recruited from classes and through advertisement on the campus of the University of Connecticut. All were native speakers of English and none had participated in a related experiment before. Subjects received either money or course credit for their participation. The paid subjects received $10 per hour. The experiment usually lasted about 30 minutes.

Materials.

Twenty experimental items were created. These are listed in Appendix 1. Each item involved four conditions as in (6):

(6) Example stimuli for Experiment 1 (A = Ambiguous; U = Unambiguous; R = Reduced; UR = Unreduced).

-6	-5	-4	-3	-2	-1		
The	coach	smiled	at	the	player...		
		0	1	2	3		
a.		tossed	a	frisbee	by...		(AR)
b.	who was	tossed	a	frisbee	by...		(AU)
c.		thrown	a	frisbee	by ...		(UR)
b.	who was	thrown	a	frisbee	by ...		(UU)

Each item included a noun phrase in a non-subject position which was modified by a relative clause in passive voice. The relative clause was either reduced (R) (6a and c) or unreduced (U) (6b and d); its verb was either ambiguous (A) (i.e. the past participle and past tense forms were homophonous and homographic—6a and b) or unambiguous (U) (6b and d).

We expected that the reduced cases would be harder than the unreduced cases for both ambiguous and unambiguous verbs. The supposition, common to the three LCAs, that local coherence causes interference independently of the difficulties associated with processing reduced relative clauses leads to the prediction of an interaction, with the greatest difficulty occurring during the processing of the relative clause in (6a).

Four counterbalanced lists of items were constructed. Each list consisted of 20 blocks of items. Each block contained 4 filler items and 1 stimulus item. The stimulus items were never first in a block. The position of the stimulus item within a block was chosen randomly, subject to the condition that the first sentence of a block had to be a filler. The first 10 items of the experiment were fillers. Each subject read six practice trials and then read one of the lists.

Procedure.

The sentences were presented on a computer monitor using non-cumulative, word-by-word, self-paced reading (Just, Carpenter, & Wooley, 1982). Each trial started with an image of the sentence in which dashes replaced all the printed characters. Participants pressed the spacebar to reveal each new word, causing the preceding word to revert to dashes. At the end of the sentence, participants answered a comprehension question by pushing "f" for Yes and "j" for No. Participants were encouraged to read as naturally as possible and to answer the questions according to their first impulse. The program recorded the responses to the questions and all response times.

The experiment was executed by a PsyScope program (Cohen, MacWhinney, Flatt, & Provost, 1993) and run on MacIntosh GE3s with 14 inch monitors. The target and control sentences were designed so that the section of the sentence from the beginning up to the fifth word beyond the ambiguous verb (or the corresponding verb in the control case) fit on one line of the computer screen.

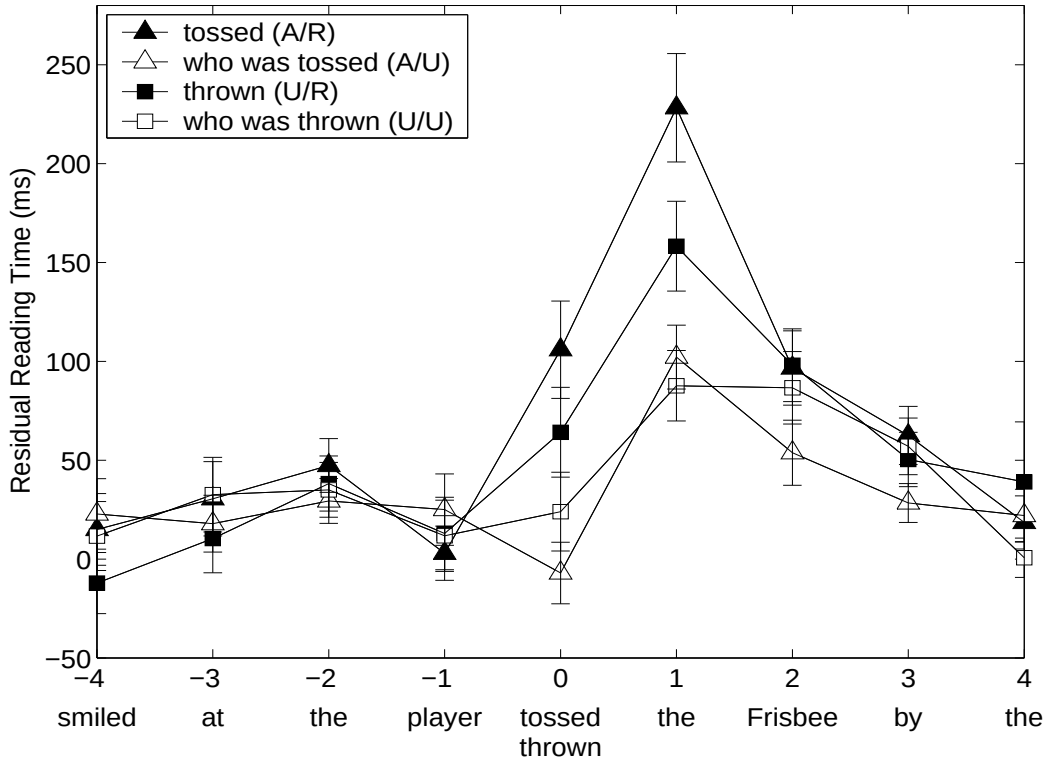
Results

The data from 9 subjects who missed more than 20% of the comprehension questions (stimuli and fillers combined) were removed from the analysis.

Before analyzing the reading times, we replaced each data point that was more than 4 standard deviations from the global mean reading time (measured over all positions and all subjects) with the value at 4 standard deviations from the mean at the position in question (global mean = 412 ms; 4 s.d.'s = 1037 ms). This adjustment of outliers affected 0.82% of the data. Likewise, for each subject, we performed a linear regression on the adjusted reading times with characters-per-word as independent variable. The variance associated with word length was again small (mean $R^2 = 0.027$), but significant ($p < 0.0001$). The analyses described below were performed on the residuals from this regression analysis.

Figure 1 shows the mean (trimmed, length-corrected) reading times across all positions for the four conditions of Experiment 1.

Figure 1. Mean residual reading times from Experiment 1. Error bars show one standard error around each data point.



We analyzed the reading times by dividing each sentence into three regions across which we expected contrast. It was at the second verb ("tossed"), that the distracting analysis was first supported by a combination of words, so we refer to this verb as the "critical word" and labelled it Word 0. We defined Region 1 to consist of Words -4 through -1. We expected a distracting effect to develop at the critical word or shortly after it, so we defined Region II to consist of Words 0 through 3; The end of the line cut some sentences off as soon as Word 5, so Region III consisted of Word 4.

For each region of interest, subject and item means were subjected to separate analyses of variance (ANOVAs), each with two factors: Ambiguity and Reduction. At Region I, there

were no significant main effects or interactions. At Region II, there was a main effect of Reduction ($F1(1, 39) = 31.03$, $MSE = 46074$, $p < .0001$; $F2(1, 19) = 33.11$, $MSE = 62680$, $p < .0001$) with slower reading times for the Reduced than the Unreduced conditions. The predicted interaction between Ambiguity and Reduction was also significant ($F1(1, 39) = 13.85$, $MSE = 46074$, $p = .0002$; $F2(1, 19) = 6.74$, $MSE = 62680$, $p = .0095$). At Region III, there were no significant main effects or interactions. We define the *Interaction Strength* as $(\overline{HR} - \overline{HU}) - (\overline{NR} - \overline{NU})$. The average interaction strengths at Regions I, II, and III, were 11, 50, and -42, respectively.

The individual significances by word position are shown in Table 1. The last line of Table 1 gives the word-by-word Interaction Strengths.

Table 1: Word-by-word significances for the Ambiguity x Reduction interaction of Experiment 1. IS = Interaction Strength = $(\overline{HR} - \overline{HU}) - (\overline{NR} - \overline{NU})$.

	-4	- 3	-2	-1	0	1	2	3	4
	smiled	at	the	player	tossed /thrown	a	frisbee	by	the
F1(1,39)	< 1	2.467	1.264	< 1	6.591*	3.320	1.874	4.389*	1.227
MSE	–	35996	23706	–	52890	58699	40456	20317	50012
p1	–	0.117	0.261	–	0.010	0.069	0.171	0.037	0.268
F2(1,19)	< 1	1.693	0.370	< 1	4.804*	0.820	0.579	2.200	2.009
MSE	–	53806	28829	–	76359	75369	58585	31971	62238
p2	–	0.194	0.543	–	0.029	0.365	0.447	0.138	0.157
IS (ms)	16	35	15	-23	73	56	31	41	-42

Note. *p < .05. **p < .01. ***p < .001.

Discussion

The results of Experiment 1 are consistent with the hypothesis that local coherence makes sentence processing difficult. In particular, they provide evidence against SCPA1, which claims that the only source of contrast between the reduced and unreduced versions of such sentences is the load associated with processing reduced relative clauses. However, there are other interpretations of these results, several of which do not require assuming that local coherence plays a role. We next consider three alternative accounts, which we label SCPA2, SCPA3, and SCPA4.

SCPA2. Since different verbs were used in the Ambiguous and Unambiguous cases, there might be a semantic/pragmatic difference between these groups that made the ambiguous cases harder. With nothing else said, such an account would not explain the interaction with Reduction. However, it is reasonable to suppose that, in a strongly constraining syntactic context (the Unreduced cases), the effect of semantic/pragmatic differences on reading times would be minimized, while in a more open-ended, or more difficult syntactic context (the Reduced cases), the effect would be amplified. SCPA2 holds that higher typicality of the intended readings in the sentences with nonhomophonous verbs were primarily responsible for the effect. To probe this hypothesis, we conducted a follow-up norming study in which we asked participants to judge the typicality of active, declarative sentences involving the relative verbs and thematic role assignments from Experiment 1 on a scale of 1 to 7.

30 participants rated examples like 7 as well as filler examples with typical and atypical role assignments (e.g., "The hunter chased the fox"; "The truck resembled a tomato.") The average ratings of the critical items (shown in parentheses in (7) differed nonsignificantly in the opposite direction to that assumed by SCPA2, casting doubt on SCPA2's claim that the main results stemmed from a typicality contrast.

(7) Someone tossed a frisbee to the player. (5.83, SE = 0.14)
Someone threw a frisbee to the player. (5.63, SE = 0.13)

SCPA3. Another possibility takes as a foundation, those grammatical theories which emphasize grammatical independence of lexical entries (e.g., Chomsky, 1981; Bresnan, 1982). Such theories are consistent with the notion that verbs vary in the degree to which they are grammatically compatible with particular syntactic environments. We did not know how to estimate the grammatical acceptability of the constructions actually tested in Experiment 1 without introducing the confound of Local Coherence. However, we noted that people's opinions seem to differ as to how felicitous it is to passivize a Recipient argument of a ditransitive verb. SCPA3 maintains that a difference in the acceptability of these passivizations is responsible for the differences observed in Experiment 1. Again, this assumption alone could not predict the observed interaction, but it could be that the strongly constraining Unreduced context minimizes these differences, while the more weakly constraining Reduced context does not. To probe SCPA3, we asked 40 participants to rate, on a scale from 1 (very unacceptable) to 7 (very acceptable) the grammatical acceptability of sentences involving Recipient and Theme extraction for both the Ambiguous and Unambiguous verbs used in Experiment 1 (8).

(8) The player was tossed a frisbee by the opposing team. (Recipient/Ambiguous) (4.30, SE = 0.15)
A frisbee was tossed to the player by the opposing team. (Theme/Ambiguous) (5.56, SE = 0.10)
The player was thrown a frisbee by the opposing team. (Recipient/Unambiguous) (4.39, SE = 0.15)
A frisbee was thrown to the player by the opposing team. (Theme/Unambiguous) (5.42, SE = 0.11)

The 20 test sentences were presented along with 40 fillers. The fillers included a number of cases which prior research has identified as ungrammatical to varying degrees, including extractions from whether-islands and extractions from subject islands (Braze, 2002; Snyder, 2000). We used cases with a range of acceptability values in order to encourage use of the intermediate values on the scale. The mean ratings are shown in parentheses in (8). An ANOVA with Ambiguity and Extraction Type as factors revealed a main effect of Extraction Type (Theme extractions rated more acceptable than Recipient extractions: $F1(39, 1) = 95.47$, $MSE = 262.205$, $p < .0001$; $F2(19, 1) = 86.78$, $MSE = 255.794$, $p < .0001$)), but no other significant effects. A scatterplot of length-corrected reading time versus acceptability of the recipient extractions indicated a linear relationship. A regression analysis indicated that reading time increased as acceptability decreased, accounting for a

small amount of variance ($R^2 = 0.0064$, $p < .0001$). Given the robust preference for Theme extraction indicated by the judgment ANOVA, one might think that it is acceptability of these extractions that is interacting with Reduction in the main experiment. But when we reran the ANOVA of Experiment 1 on the residuals from the acceptability regression, all the same contrasts were significant. This result casts some doubt on SCPA3's claim.

SCPA4. Swinney (1979) provided evidence that when a semantically ambiguous word is read in a constraining context, both meanings are briefly activated. Tanenhaus, Leiman, and Seidenberg (1979) provided evidence that the same was true for syntactically ambiguous words. Gibson and Tunstall (1999) argued that competition from incorrect senses of ambiguous words might be responsible for some observed cases of difficulty in sentence processing. These observations motivate SCPA4, which assumes that the syntactic requirements of the current context have to be reconciled with the lexically-specified properties of each word as it is read; when the bias of the context is at odds with the bias of the word, reading is slowed. Such an approach could explain why the ambiguous verbs of Experiment 1, which are plausibly biased toward their active interpretations, were harder to process than the unambiguous verbs, which may be biased toward the passive interpretation which the syntactic context requires.[1] Probabilistic grammar models which take into account both structural and lexical biases (e.g., Crocker & Brants, 2000; Jurafsky, 1996a) are well-positioned to make predictions along these lines. Such an account does not immediately explain the interaction between Ambiguity and Reduction, but once again, one might suppose that a more strongly constraining syntactic context (the Unreduced context) minimizes interference from irrelevant lexical senses.

Experiment 2: A Semantic Control

If SCPA4 is correct, then there is no need to posit the construction of locally-motivated syntactic structures that are at odds with the global parse. In essence SCPA4 treats the interaction of Experiment 1 as stemming from a kind of conflict that is naturally predicted by grammatical architectures that separate syntax and lexicon. Thus, to further probe the claim that merely locally coherent syntactic structures exert an influence, Experiment 2 used a different design in which the same verb was used in all four conditions. If the same pattern of results occurs in a case in which there is no comparison between verbs, then we can rule out explanations which refer to lexical valence contrasts.

Experiment 2 was also motivated by a contrast with the literature on self-organizing models. Vosse and Kempen (2000)'s model posits only syntactic features on the attachment sites of the self-organizing tree-fragments. Tabor and Hutchins (in press) claims that semantic as well as syntactic features guide the parsing. If the latter claim is correct, then a local structure with both syntactic and semantic coherence ought to produce a more pronounced distraction effect than a local structure with only syntactic coherence. Experiment 2 also tests this prediction.

[1] This explanation is complicated by the fact that lexically, even what we have called the "unambiguous" verbs are still open to several syntactic interpretations. For example, "thrown" can occur in Recipient-extracted sense intended here, or in a Theme-extracted passive sense ("The frisbee was thrown to the player") or in an active sense ("The player has thrown the frisbee"). SCPA4 assumes that the Recipient-extracted passive sense is preferred for "thrown" but the active sense is preferred for "tossed".

Method

Participants.

Eighty-six subjects were recruited from classes on the campus of the University of Connecticut. All were native speakers of English and none had participated in a related experiment before. The participants received course credit for their participation. The experiment usually lasted about 30 minutes.

Materials.

We considered reusing the Ambiguous stimuli from Experiment 1 and creating controls that lacked semantic coherence in the critical region. However, we found that it was hard to design felicitous sentences with substantial variation in semantic plausibility under this rubric because the constraints on the semantic attributes of a ditransitive Recipient are quite strict. Therefore, we created an alternative design by replacing the ditransitive ambiguous verbs with transitive verbs. To take advantage of semantic biases which have already been experimentally measured, we used a subset of the noun phrase + transitive verb combinations employed by Trueswell, Tanenhaus, and Garnsey (1994), a study which examined the effect of noun phrase animacy on the processing of standard Main Verb/Reduced Relative ambiguities. We started by creating contrast pairs like (9a-b).

(9) a. The bandit worried about the prisoner transported by the capricious guards.
b. The bandit worried about the gold transported by the capricious guards.

Although we expected a distracting effect of semantic cohesion for a simple sequence like "the prisoner transported" (versus "the gold transported"), we thought an even longer local ambiguity might produce a more pronounced effect, because there would be more words supporting the formation of the local structure. Therefore, we added time or path adverbials of the form "the" + Adjective + Noun (e.g., "the first time", "the whole way") after each ambiguous verb. To control for potential reading time differences stemming from lexical contrasts between the manipulated nouns, we included a reduction manipulation like that of Experiment 1. An example item from the resulting 2 x 2 (Animacy x Reduction) design is shown in (10). Appendix 2 lists the stimuli for Experiment 2. Four counterbalanced lists of items were created, as in Experiment 1.

(10) Example stimuli for Experiment 2. (A = Animate; I = Inanimate; R = Reduced; U = Unreduced)

The bandit worried about ...

	-2	-1			0	1	2	3	4	
a.	the	prisoner			transported	the	whole	way	by	(A / R)
a.	the	prisoner	who	was	transported	the	whole	way	by	(A / U)
a.	the	gold			transported	the	whole	way	by	(I / R)
a.	the	gold	that	was	transported	the	whole	way	by	(I / U)

... the capricious guards.

Procedure.

The procedure was the same as in Experiment 1 except that some of the participants were tested in groups of up to 8 at a time. They sat at separate computer terminals, spaced at least 3 feet apart from one another and wore headphones to block out ambient noise.

Results

The data from fifteen subjects who missed more than 20% of the comprehension questions (stimuli and fillers combined) and from one subject whose computer crashed partway through the experiment were removed from the analysis.[2]

As in Experiment 1, before analyzing the reading times, we replaced each data point that was more than 4 standard deviations from the global mean reading time (measured over all positions and all subjects) with the value at 4 standard deviations from the mean at the position in question (mean = 408 ms; 4 s.d.'s = 1005 ms). This adjustment of outliers affected 0.85% of the data. Likewise, for each subject, we performed a linear regression on the adjusted reading times with characters-per-word as independent variable. The variance associated with word length was again small (mean $R^2 = 0.016$), but significant ($p < 0.0001$). The analyses described below were performed on the residuals from this regression analysis.

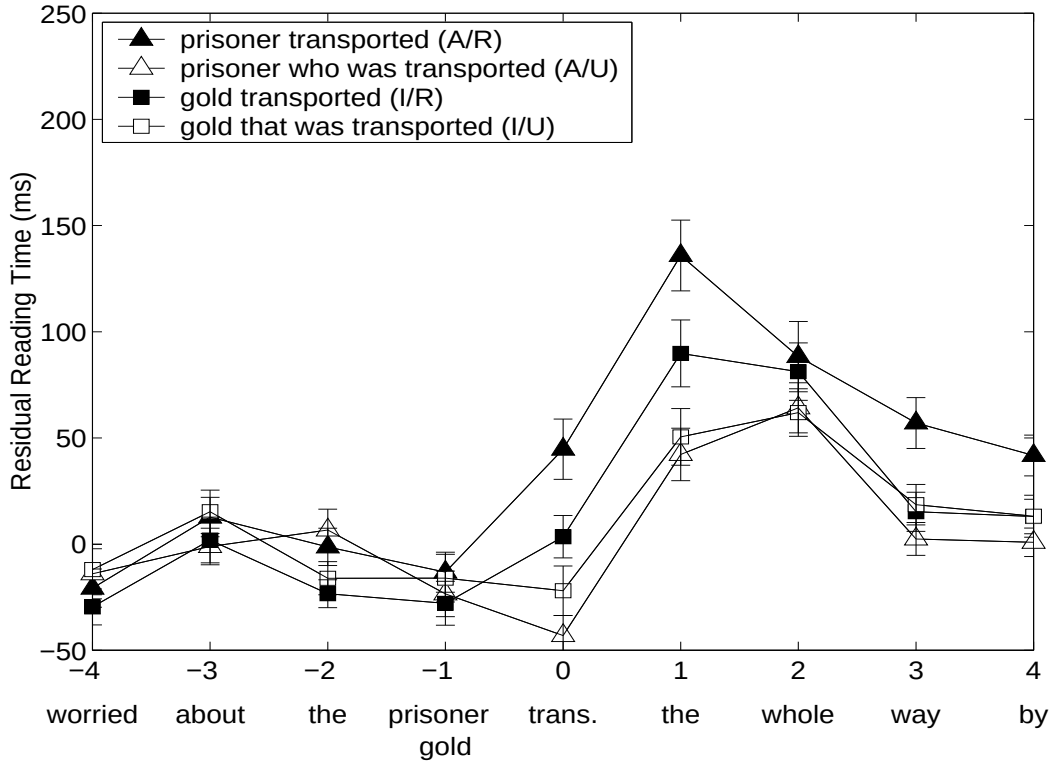
Figure 2 shows the mean (trimmed, length-corrected) reading times across positions for the four conditions of Experiment 2.

As in Experiment 1, we analyzed the reading times by dividing each sentence into three regions across which we expected contrast. Again, it is at the second verb (here, "transported"), that the distracting analysis is first supported by a combination of words, so the second verb was defined as Word 0. Region I consisted of Words -4 through -1; Region II consisted of Words 0 through 3; Region III consisted of Word 4 (again, the last word on the line was sometimes Word 5).

For each region of interest, subject and item means were subjected to separate analyses of variance (ANOVAs), each with two factors: Animacy and Reduction. At Region I, there were no significant main effects or interactions. At Region II, there was a main effect of Reduction ($F1(1, 69) = 44.336$, $MSE = 2865$, $p < .001$; $F2(1, 19) = 10.512$, $MSE = 2873$, $p = .004$). The predicted interaction between Animacy and Reduction was significant ($F1(1, 69) = 15.688$, $MSE = 2245$, $p < .001$; $F2(1, 19) = 7.043$, $MSE = 2522$, $p = .016$). At Region III, the main effect of Reduction was significant in the Subjects analysis and marginally significant in the Items analysis ($F1(1, 69) = 6.455$, $MSE = 4494$, $p = .013$; $F2(1, 19) = 4.152$, $MSE = 2821$, $p = .056$). The Region III interaction between Animacy and Reduction was significant in the Subjects analysis only ($F1(1, 69) = 6.229$, $MSE = 4731$, $p = .015$; $F2(1, 19) < 1$). The individual significances by word position are shown in Table 2. The average Interaction Strengths by region were: Region I—15 ms/word; Region II—45 ms/word; Region III—41 ms/word. The individual tests showed one small (23 ms) but significant main effect at a position prior to the manipulation: the effect of Animacy at Position -2 ($F1(1, 69) = 6.518$, $MSE = 5372$, $p = .013$; $F2(1, 19) = 20.300$, $MSE = 835$, $p < .001$).

[2]Due to an error in the implementation, the distribution of trials over conditions for two items (16 and 17) was strongly asymmetric. Nevertheless, the results of the analysis were not significantly different when these two items were removed. Therefore, they are included in the analyses reported below.

Figure 2. Mean residual reading times from Experiment 2. Error bars show one standard error around each data point.



Discussion.

These results are not consistent with SCPA4, the account which claims that a conflict between the lexical bias of the critical verb and the syntactic properties of its environment is responsible for the effects. In this case, the same critical verb is used in all four conditions and the only syntactic manipulation is the Reduction contrast, so such an account cannot explain the observed interaction. The Local Coherence Accounts do predict this interaction on the assumption that semantic coherence increases local coherence.

Corley and Crocker (2000) propose a model which is similar in conception to SCPA4: a tagging module assigns part-of-speech tags to words; these are submitted to a syntactic module for parsing; if the syntactic module cannot parse what the tagging module gives it, then reanalysis must occur, and reading slows down. The tagger chooses the most likely tag for each word, taking into account the likelihood of the word given the tag (a unigram statistic) and the likelihood of the tag given the previous tag (a bigram statistic—see Charniak, 1993). Depending on the design of the tag set and the probabilities involved, this model might predict reanalysis only in the Reduced/Ambiguous case of Experiment 1, thus predicting an interaction like the one we observed in that experiment. Such a model would not, as formulated, predict the interaction observed at the second verb (Word 0) in Experiment 2 because the tagging does not take into account the identity of the preceding word. One might suggest changing it to let such lexically specific information influence the tagging.

Table 2: Word-by-word significances for the Animacy x Reduction interaction of Experiment 2. IS = Interaction Strength = $(\overline{AR} - \overline{AU}) - (\overline{IR} - \overline{IU})$.

	-4 worried	-3 about	-2 the	-1 prisoner /gold	0 transported	1 the	2 whole	3 way	4 by
F1(1, 69)	< 1	2.254	< 1	1.764	13.761***	4.777*	< 1	10.917**	6.229*
MSE	–	5721	–	4822	4945	10821	–	5382	4731
p1	–	.138	–	.188	$< .001$	.032	–	.002	.015
F2(1, 19)	< 1	1.408	< 1	2.355	5.983*	3.847	< 1	9.488**	< 1
MSE	–	2385	–	3870	4318	6711	–	5527	–
p2	–	.250	–	.141	0.024	.065	–	0.006	–
IS (ms)	11	27	-1	22	62	54	5	58	41

Note. *p < .05. **p < .01. ***p < .001.

Likewise, one might predict local coherence effects over longer intervals by using n-grams with larger n's (see Crocker & Brants, 2000). However, despite its superficial resemblance to SCPA4, if such an account turns out to be the correct theory of parsing, our claim that locally coherent syntactic structures can distract the parser will be vindicated: the use of n-grams in the lexical module gives the model limited sensitivity to local coherence. Thus we view this approach as a kind of Local Coherence account. It is a member of the class we called "LCA2" above.

It is also possible that differential typicality of the events described by the intended grammatical reading is responsible for the contrast between the Animate and Inanimate conditions in Experiment 2—the hypothesis we referred to as "SCPA2" above. Again, such an effect might plausibly be amplified in the Reduced conditions and minimized in the Unreduced conditions to produce the observed interaction. To test the plausibility of SCPA2 for Experiment 2, we included in the norming study described above examples like (11), which are derived from the Experiment 2 stimuli.

(11) The prisoner was transported. (Animate/Passive) (6.06, $SE = 0.12$)
The prisoner transported something. (Animate/Active) (6.03, $SE = 0.12$)
The gold was transported. (Inanimate/Passive) (6.11, $SE = 0.12$)
The gold transported something. (Inanimate/Active) (2.16, $SE = 0.14$)

Recall that in that norming study, participants were asked to judge, on a scale from 1 to 7, the typicality of the event described by each sentence. The average judgments are shown in parenthesis in (11). The interaction between Animacy and Voice was significant, and the difference between the Inanimate/Active condition and the other three conditions were significant according to post hoc tests (Tukey's). Numerically, the Inanimate/Passive condition was rated slightly more typical than the Animate/Passive condition (6.11 vs. 6.06) in keeping with SCPA2's claim that a contrast in the typicality of these role assignments was responsible for the observed effect, but this effect was small and nonsignificant. In fact, no other pairwise comparisons were significant in post hoc tests. To test the possibility that a contrast in the typicality of the Passive conditions was the underlying factor that interacted with Reduction in Experiment 2, we computed residuals from a regression of

length-corrected reading time versus typicality and reran the main ANOVA. All of the same effects were significant, casting doubt on SCPA2's claim that a typicality difference between the two passivizations led to the critical interaction.

SCPA5. There is one other kind of self-consistent parse account which should mentioned. It may be that some of the stimuli are more grammatical than others; hence grammar, not processing, is ultimately responsible for the contrasting reading times observed in Experiments 1 and 2. We have already provided evidence against one version of this approach: the version that claims that Recipient-argument extraction is less acceptable for the Ambiguous verbs than the Unambiguous verbs of Experiment 1. Another possibility is simply that the Ambiguous/Animate Reduced conditions of both experiments are less grammatical than the other conditions, either absolutely, or under a theory which posits graded grammaticality values. We have not tested this claim in the present studies for two reasons: (1) We know of no linguistic motivation for setting the Ambiguous/Animate Reduced conditions apart—indeed reduced and unreduced relative clauses appear in virtually the same syntactic environments, and we know of no syntactic rule of English that is sensitive to either ambiguity or animacy contrasts. (2) We do not know how to directly assess the grammaticality of the critical condition in the laboratory without introducing the confound of Local Coherence. If a relevant grammatical distinction is identified, then SCPA5 will be worth pursuing further.

Interpretation of Locally Coherent Structures.

A further prediction distinguishes the Local Coherence Accounts from the Self Consistent Parse Accounts: if we assume that once the locally coherent structures are constructed, they are interpreted, then there should be a comprehension difference that correlates with the presence of local coherence.

We ran a follow-up self-paced reading study on shortened versions of the the two Animate conditions of Experiment 2 (Unreduced and Reduced) to address this issue. We asked one of two questions after each condition, as shown in (12).

(12) The bandit worried about the prisoner (who was) transported the whole way.
  a. Was the prisoner transported? (prisoner = Theme of transporting)
  b. Did the prisoner transport something? (prisoner = Agent of transporting)

We omitted the material from the Agent phrase onward (e.g., "by the guards...") because we wanted to find out what people were understanding immediately after reading the segment in which the reading time differences appeared. The filler:stimulus ratio was 3:1. The same procedure was used as in Experiments 1 and 2. Thirty-eight undergraduate native English speakers from the University of Connecticut participated in the study for course credit. The Local Coherence Accounts predict an interaction between Reduction and Question type: readers should be more likely to consider the ambiguous noun phrase ("prisoner") to be an Agent and less likely to consider it to be a Theme in the Reduced condition than the Unreduced.

The data from two participants, who made more than 25% incorrect responses on the filler comprehension questions were removed from the analysis. The results are shown in Table 3. Subject and item means were subjected to separate analyses of variance (ANOVAs), each with two factors: Reduction and Question Type. There was a main effect of Question

Table 3: Mean rate of "Yes" responses in the four conditions of the Experiment 2 followup study. "Theme/Agent question" means that the question asked if the ambiguous noun phrase was a Theme/Agent of the subsequent verb, respectively. The numbers in parentheses are standard errors.

	Theme question	Agent question
Unreduced	0.96 (0.013)	0.18 (0.034)
Reduced	0.90 (0.021)	0.27 (0.038)

Type ($F1(1, 35) = 535.228$, $MSE = 0.0335$, $p < .001$; $F2(1, 19) = 341.093$, $MSE = 0.0292$, $p < .001$). The predicted interaction between Question Type and Reduction was significant ($F1(1, 35) = 9.820$, $MSE = 0.0188$, $p = .003$; $F2(1, 19) = 6.057$, $MSE = 0.0170$, $p = .024$), although the effect of Reduction was only significant for the Theme questions in a post hoc test (Fisher's Protected Least Squares, Theme question: $p = .013$; Agent question: $p = .117$). There was no main effect of Reduction. T-tests revealed that the total rate of "Yes" responses (Agent and Theme responses added together) was greater than 100% in both the Reduced and the Unreduced sentences.

Although the effect is not large, the tendency toward a greater ratio of Agent assertions to Theme assertions in the Reduced condition is consistent with the claim of the Local Coherence Accounts that the erroneous parse becomes at least temporarily established in the Reduced condition.

Thus far, we have found that several plausible self-consistent approaches to the phenomenon at hand are not supported by evidence that we have been able to discern. This does not mean there is no self-consistent explanation, for we may not yet have applied the correct test. We now turn our attention to the matter of the grammaticality of the stimuli, a point highlighted by LCA3.

Experiment 3: Grammaticality

Experiment 3 takes a step toward assessing the role of grammaticality in determining our findings, particularly in light of the fact that Experiment 1 involves a fairly obscure construction—Recipient extraction from ditransitive verbs—which some speakers may not find grammatical. Recall that LCA3 claims that local coherence effects stem from parse failure, followed by a salvaging operation which attempts to build a parse of part of the input and succeeds in doing so only in the locally coherent conditions.

In order to assess how participants from our subject pool judged grammaticality under the conditions in which the original experiments were run, we reran the self-paced word-by-word reading and asked for grammaticality judgments instead of answers to comprehension questions. We asked for yes/no grammaticality judgments on the assumption that this would allow us to straightforwardly assess the possibility that the Experiment 1 Recipient extractions were ungrammatical in an absolute sense (finding contexts in which subjects reliably classified this construction grammatical would argue against an absolute ungrammaticality interpretation).

Although asking naive participants for end-of-sentence grammaticality ratings may not succeed in revealing whether parse failure plays a role in the processing (the parse failure might not be consciously detectable at the point of the query), low rates of positive grammaticality judgment could be an indication that the stimuli regularly foment parse

failure, thus supporting LCA3's premise.

We also hoped to replicate the critical interaction observed in the first two experiments using a different behavioral measure.

Participants.

Twenty-eight subjects were recruited by advertisement on the campus of the University of Connecticut. All were native speakers of English and none had participated in a related experiment before. They were paid at a rate of $10 per hour for their participation in the experiment. The experiment usually lasted about 30 minutes.

Materials.

Experiment 3 was a composite of two subexperiments, 3.1 and 3.2. The stimuli for these subexperiments were shortened versions of the stimuli from Experiments 1 and 2, respectively. Each stimulus ended after the agentive by-phrase (as indicated in Appendices 1 and 2), because we wanted to measure grammaticality judgments as soon as possible following the critical region (Region II) of Experiments 1 and 2. Four lists of items were constructed. Each list included 40 blocks of items, 20 from Experiment 1 and 20 from Experiment 2. Each block contained 2 filler items and 1 stimulus item. The stimulus items were never first in a block. Half of the filler sentences were grammatical and half were ungrammatical. The ungrammatical filler sentences were phrase-structurally defective strings which could be construed as being related to grammatical sentences in the sense that they lacked a critical constituent (e.g., "The girls consumed the grapefruit which the cat.") or in that they contained an unlicensed constituent (e.g., "The rotten mean uncle the teacher graded the children's work."). The ungrammatical stimuli were a slightly modified subset of the ungrammatical stimuli used by Ferreira and Henderson (1991) in a grammaticality judgment study. The position of the stimulus item within a block was chosen randomly, subject to the condition that the first sentence of a block had to be a filler. Each subject read six practice trials and then read one of the ($3 \times 40 =$) 120-trial lists.

We predicted that in each subexperiment, the positive grammaticality judgment rate would interact in the same way as the reading times did in the corresponding reading experiments.

Procedure.

The procedure was the same as in Experiments 1 and 2 except that all participants were run in groups of up to 8 at a time (spaced 3 feet apart, wearing headphones), each sentence was followed by the question,"Grammatical? f: Yes j: No", and no feedback was provided. We culled only the question response data from the output of the program.

Results

No subject disagreed with our judgments on more 20% of the filler trials and all of the data were used in the analysis. Table 4 shows the mean values of the grammaticality judgments for each condition and subexperiment.

The two subexperiments were analyzed separately. Subject and item means were subjected to separate analyses of variance (ANOVAs). The strengths of the main effects and interactions are shown in Table 5. As the tables indicate, Reduction has the effect of reducing the rate of positive grammaticality judgment in both subexperiments and it reduces the

Table 4: Mean rates of positive grammaticality judgments for the two subexperiments and the fillers in Experiment 3. The values shown in parentheses are standard errors.

(a) Subexperiment 3.1.

Ambiguous/Reduced	0.22 (0.040)
Ambiguous/Unreduced	0.72 (0.033)
Unambiguous/Reduced	0.39 (0.042)
Unambiguous/Unreduced	0.72 (0.038)

(b) Subexperiment 3.2.

Animate/Reduced	0.58 (0.045)
Animate/Unreduced	0.82 (0.030)
Inanimate/Reduced	0.75 (0.035)
Inanimate/Unreduced	0.84 (0.032)

(c) Fillers

Grammatical	0.93 (0.012)
Ungrammatical	0.08 (0.013)

rate more in the Ambiguous/Animate conditions than the Unambiguous/Inanimate conditions. In particular, the predicted interaction between Ambiguity/Animacy and Reduction was significant in both experiments.

Discussion.

The form of the interactions in this experiment replicates the form of the interactions in the previous two experiments, thus indicating that the phenomenon under study is detectable via two distinct experimental measures. All three of the Local Coherence Accounts predict the grammaticality judgment results with the addition of some reasonable assumptions. On the assumption that the effort associated with building a partial parse after the global parse has failed produces a higher rate of negative grammaticality judgment, LCA3 predicts the particularly low judgments in the Ambiguous/Animate, Reduced conditions. LCA2 can predict the results in a similar fashion: by claiming that conflict between the tagger and the parser leads to low rates of positive grammaticality judgment. Under the assumption that there is some noise in the activations of lexical tree-fragments (Vosse & Kempen, 2000), the self-organization approach (LCA1) also predicts elevated rates of parse failure in the same cases. This is because the global parse is more likely to be stymied by errant local structure building in the cases (Ambiguous/Animate, Reduced) where the local structure is more robust.

Experiment 3 provides empirical evidence against the claim that Recipient and Theme Extractions in relative clauses are ungrammatical in an absolute sense. The Unreduced versions of each sentence were judged grammatical at a relatively high average rate and 18 different Unreduced items (9 from Experiment 1 and 9 from Experiment 2) were judged grammatical 100% of the time (9 out of 9 trials). Given the low rate at which uncontroversially ungrammatical constructions were accepted (8%), these results would be highly

Table 5: Experiment 3 ANOVA results.

(a). Subexperiment 3.1: Ambiguity and Reduction.

	Subjects			Items		
Effect	$F(1, 35)$	*MSE*	p	$F(1, 19)$	*MSE*	p
Ambiguity	12.096**	0.0221	.001	3.776	0.0687	.067
Reduction	148.078***	0.0428	$< .001$	97.303***	0.0371	$< .001$
Ambiguity × Reduction	7.321*	0.0341	.01	5.668*	0.0229	.028

(b). Subexperiment 3.2: Animacy Reduction.

	Subjects			Items		
Effect	$F(1, 35)$	*MSE*	p	$F(1, 19)$	*MSE*	p
Ambiguity	12.364***	0.0252	.001	7.946*	0.0224	.011
Reduction	29.238***	0.0325	$< .001$	13.977***	0.0322	$< .001$
Ambiguity × Reduction	5.852*	0.0404	.021	4.785*	0.0218	.041

Note. *p < .05. **p < .01. ***p < .001.

unlikely if the underlying syntactic constructions were absolutely ungrammatical for our participants. Thus the Experiment 3 findings cast doubt on the hypothesis that the cause of the parse failure posited by LCA3 is simple unconstructability of the parses for all the stimuli. A more plausible viewpoint would seem to be that parsing fails at some rate for processing reasons (perhaps there is some noise in the parse building mechanisms), and this failure opens the door to parse salvaging.

On the other hand, the particularly low rates of positive grammaticality judgments in the experimental conditions compared to the grammatical controls is consistent with LCA3's claim that the effects of local coherence are a phenomenon of ungrammatical sentence processing. LCA1 and LCA2 are compatible with the low rates of positive judgment, but they do not require them. The low overall rates of positive grammaticality judgment also raise the question as to whether self-paced reading provides an accurate indication of how sentence processing works under normal reading conditions. It will be valuable, in future work, to determine whether the pattern revealed here occurs under more typical reading conditions.

GENERAL DISCUSSION

We have reported on two word-by-word self-paced reading experiments and two grammaticality judgment experiments which probed for effects of local syntactic coherence in the input. All four experiments exhibited an interaction between Reduction and the presence versus absence of a (merely) locally coherent active clause.

We considered several explanations for these effects which treated the presence of locally coherent structure as incidental and sought explanations for the interaction in possible differences among the self-consistent parses. We referred to these explanations as Self-Consistent Parse Accounts (SCPAs). Followup norming studies which probed for semantic and syntactic differences between the stimuli did not find evidence to support the

assumptions of the particular SCPAs we considered.

We also considered explanations which posit that, at some stage of processing, the parser parses the locally coherent structure. Local Coherence Account 1 (LCA3) claims that local coherence only has an influence after an attempt to find a globally coherent structure has failed. By contrast, LCA1 and LCA2 claim that the presence of mere local coherence can interfere with the construction of a globally consistent parse.

Related Work

Several other studies have provided evidence that is relevant to the plausibility the self-organization view (LCA1).

Related Work: Reanalysis as a Last Resort.

Following Fodor and Frazier (1980)'s proposal that the parser prefers to avoid reanalysis whenever other options or possible, Schneider and Phillips (2001) and Sturt, Pickering, Scheepers, and Crocker (2001) set up experiments in which they pitted easy reanalyses against other constraints. Their conclusion was that the parser generally prefers to keep structure that it has already built. It is worth noting that the self-organization approach makes a similar, general prediction: since parses are preferred to the extent that they are supported by the evidence of the input, and since an initial analysis gets a head start on becoming activated compared to a competing analysis that is first signalled by late-arriving disambiguating information, the initial analysis has an advantage (see Tabor & Hutchins, in press). Nevertheless, we are claiming that the current experiments involve cases in which the later-arriving information has the capacity to at least strongly interfere with the initial analysis. In fact, Schneider and Phillips (2001)'s and Sturt et al. (2001)'s stimuli (illustrated in (13) and (14) respectively) afford local analyses like ours, and in their cases, these local analyses are even consistent with a grammatical parse of the entire input. But in both cases, the researchers found that the "himself" version was dispreferred, indicating that if the locally coherent syntactic structure had any cohesion tendency, it was not strong enough to decide the bias in the reading preference in these cases. These results thus seem, at first glance, to cast doubt on the self-organization approach.

(13) The woman who knows the funny man wrote some comedy sketches himself/herself... (from Schneider & Phillips, 2001).

(14) The troops who found the enemy spy had shot himself/themselves... (from Sturt et al., 2001).

However, these cases are not necessarily inconsistent with self-organized parsing. Both the high and low attachment interpretations of Schneider and Phillip's study and of Sturt et al.'s study involved the same preferred, active interpretations of the critical second verb. By contrast, in our experiments, the locally coherent structure involves the preferred, active reading of the critical verb while the grammatically coherent reading involves a dispreferred valence. A second difference between the experiments stems from the fact that in the preferred reading of the Schneider and Phillips and Sturt et al. stimuli, the agent of the critical verb has already been overtly identified (by the first noun phrase of the sentence) when the critical verb is read, and indeed, has generated an unsatisfied dependency in

the sense of Gibson and Thomas (1996) and Gibson (1998) which the critical verb can satisfy. By contrast, in the grammatically coherent reading of our stimuli, the agent of the troublesome verb is an element whose existence has to be inferred at the point of reading the verb. These differences may have caused readers to prefer building the locally coherent structures in our experiments and to disprefer doing so in Phillips and Gibson's and Sturt et al.'s. Note, however, that the Phillips and Gibson and Sturt et al. comparisons were not designed to address the question of whether local coherence was playing a role. To address this question, one would want to compare cases in which the local sequence had more or less coherence and see what effect this contrast had on the preferences for pronominal reference.

Related Work: Good Enough Parsing.

The results of the followup study to Experiment 2 on interpretation of locally coherent structures bear a resemblance to the findings of Christianson, Hollingworth, Halliwell, and Ferreira (2001) and Ferreira, Christianson, and Hollingworth (2001) who studied people's comprehension of sentences like (15) by examining their answers to comprehension questions presented after the sentences had been read.

(15) While Anna dressed the baby spit up on the bed.

They found that people had a significant tendency to conclude, in such examples, that Anna dressed the baby as well as that the baby spit up. They attribute their result to the formation of "Good Enough Representations", or syntactic representations that have some loose ends but are good enough to get the job done.

The Local Coherence Accounts provide another way of understanding these cases: readers may construct both interpretations, at least temporarily. However, it is also possible that asking the comprehension question activates the interpretation that the question is about, leading to the overall yes-response bias that we observed.

Conclusion

One of the main sources of interest of the present empirical results is that they draw attention to the distinctions among LCA1, LCA2, and LCA3. One issue, in particular, is subtle. We need to know whether the formation of a locally coherent structure is a cause of initial parse disturbance (LCA1 and LCA2) or a consequence of it (LCA3). Since we do not feel that our data clearly resolve this issue, we will leave this as an open question. It is worth noting that all three approaches require computational mechanisms which have not yet been thoroughly developed. LCA3 requires a mechanism for figuring out how to construct a partial parse from input for which a global parse has been rejected. This raises the question of how the parser decides which parts of the input to include in a partial parse. LCA2 involves deciding exactly how much syntactic knowledge to make available to a part-of-speech tagger. LCA1 requires the design of a set of word- or morpheme-based constraints and a set of governing dynamical equations which can parse effectively to the extent that people do. Future computational studies may help resolve the question of which mechanism is more viable as a computational model. Future empirical studies may help resolve the cause/effect question by finding out whether local coherence exerts an influence in circumstances where parse failure is not plausible. In conjunction with the data at hand,

all three viewpoints motivate looking more carefully into the question of how the parser handles grammatical inconsistency.

References

Braze, F. (2002). *Grammaticality, acceptability, and sentence processing: A psycholinguistic study.* (Unpublished PhD Dissertation, University of Connecticut)

Bresnan, J. (1982). The passive in lexical theory. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 3–86). Cambridge, MA: MIT Press.

Charniak, E. (1993). *Statistical language learning.* Cambridge, Massachusetts: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

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

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257–271.

Corley, S., & Crocker, M. W. (2000). The modular statistical hypothesis: Exploring lexical category ambiguity. In M. W. Crocker, M. Pickering, & J. Charles Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 135–160). Cambridge: Cambridge University Press.

Crocker, M., & Brants, T. (2000). Wide coverage probabilistic sentence processing. *Journal of Psycholinguistic Research*, *29*(6), 647–669.

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*(1), 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. (1990). The use of verb information in syntactic parsing: A comparison of evidence from eye movements and segment-by-segment self-paced reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 555–568.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745.

Fodor, J. D., & Frazier, L. (1980). Is the human sentence parsing mechanism an atn? *Cognition*, *8*, 417–459.

Frazier, L. (1978). *On comprehending sentences: syntactic parsing strategies.* (Unpublished doctoral dissertation, University of Connecticut)

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance* (p. 559-586). Hillsdale, NJ: Lawrence Earlbaum Associates.

Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge, Massachusetts: MIT Press.

Frazier, L., & Clifton, C. (1998). Sentence reanalysis and visibility. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 143–176). Dordrecht: Kluwer.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*(6), 291-325.

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

Gibson, E. (1991). *A computational theory of human linguistic processing. memory limitions and processing breakdown.* (Ph.D. Dissertation, Carnegie Mellon University)

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E., & Thomas, J. (1996). The processing complexity of english center-embedded and self-embedded structures. In C. Schütze (Ed.), *Proceedings of the NELS 26 workshop on language processing.* Cambridge, MA: MIT occasional papers in linguistics 9.

Gibson, E., & Tunstall, S. (1999). *The parser need not keep track of contingent frequencies: a lexical account of the english determiner/complementizer ambiguity.* (Poster session presented at the 12th Annual CUNY Sentence Processing Conference, New York, New York)

Gorrell, P. (1987). *Studies in human syntactic processing: Ranked-parallel versus serial models.* (Unpublished doctoral dissertation, University of Connecticut)

Gorrell, P. (1995). *Syntax and parsing.* Cambridge: Cambridge University Press.

Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language*, *46*, 267–295.

Hale, J. (2002). *The information conveyed by words in sentences.* (Paper presented at the 15th Annual CUNY Conference on Human Sentence Processing, New York, New York)

Jurafsky, D. (1996a). Conditional probabilities for linguistic access and disambiguation. (Paper presented at the 1996 CUNY Sentence Processing Conference, New York, New York.)

Jurafsky, D. (1996b). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.

Just, M. A., Carpenter, P. A., & Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228–238.

Kempen, G., & Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: a cognitive architecture based on activation decay and simulated annealing. *Connection Science*, *1*, 273–290.

MacDonald, M. A., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in online sentence comprehension. *Journal of Memory and Language*, *38*, 283-312.

Phillips, C. (1998). *Linear order and constituency.* (Manuscript, University of Maryland; revised version to appear in Linguistic Inquiry)

Schneider, D., & Phillips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language*, *45*, 308–336.

Snyder, W. (2000). An experimental investigation of grammaticality judgments. *Linguistic Inquiry*, *31*(3/4).

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, *23*(4), 295–322.

Stevenson, S. (1997). An integrated symbolic/connectionist parsing archicture. In R. Sun & F. Alexandre (Eds.), *Connectionist-symbolic integration: From unified to hybrid approaches* (pp. 209–206). Mahwah, NJ: Lawrence Erlbaum Assoc.

Stevenson, S. (1998). Parsing as incremental restructuring. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 327–363). Dordrecht: Kluwer Academic Publishers.

Stevenson, S., & Merlo, P. (1997). Lexical structure and parsing complexity. *Language and Cognitive Processes*, *12*(2/3), 349–399.

Sturt, P., Pickering, M. J., Scheepers, C., & Crocker, M. W. (2001). The preservation of structure in language comprehension: Is reanalysis the last resort? *Journal of Memory and Language*, *45*, 283–307.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–660.

Tabor, W., & Hutchins, S. (in press). Evidence for self-organized sentence processing: Digging in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Tabor, W., Juliano, C., & Tanenhaus, M. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, *12*(2/3), 211–271.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, *18*, 427–440.

Traxler, M. J., Pickering, M. J., & C. Clifton, J. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*(558–592).

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

vanGompel, R., Pickering, M., & Traxler, M. (2001). Reanalysis in sentence processing: evidence against current contraint-based and two-stage models. *Journal of Memory and Language*, *45*, 225–258.

vanGompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648). Oxford: Elsevier.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105–143.

Appendix 1. Stimuli for Experiments 1 and 3

The vertical slash (|) indicates the location of the end of the sentence in Experiment 3.

1. The kindergartners liked the little girl (who was) brought/chosen a toy by her parents | on the first day of Chanukah.
2. We saw a movie about an artist (who was) painted/drawn a picture by her father | while he was on his deathbed.
3. The health officials pounced on a restaurant (which was) sent/flown a shipment of salmon by a company | that had failed to comply with refrigeration laws.
4. At the dinner party, I met a man (who was) allowed/forbidden the pleasure of eating sweets by his doctor. |
5. One expects a man (who is) told/forgiven his sins by his own god to have tolerance for weaknesses in others.
(Experiment 3: One should respect a man told/forgiven his sins by his own god.)
6. An elderly gentleman addressed the woman (who was) offered/given a beer by the hostess. |
7. Balthazar praised the professor (who was) taught/given Swahili lessons by a graduate student. |
8. The manager watched a waiter (who was) served/given pea soup by a trainee. |
9. James entertained the children (who were) dyed/hidden Easter eggs by their teachers. |
10. The foreman yelled at a carpenter (who was) cut/sawn a board by his buddy. |
11. The preschool teacher congratulated the little boy (who was) knitted/sewn a hat by his grandmother. |
12. The janitor chatted with the young man (who was) rented/shown an apartment by his uncle. |
13. The nurse admonished a student (who was) nabbed/stolen a muffin by her friends | from the dining hall.
14. The play centered around an innkeeper (who was) recited/sung a verse by a travelling monk. |
15. The hotel owner questioned a guest (who was) brought/taken a drink by the bellboy. |
16. The coach smiled at the player (who was) tossed/thrown a frisbee by the opposing team. |
17. The anthropologist interviewed a woman (who was) knitted/woven a shawl by her mother. |
18. The FBI questioned a congressman (who was) mailed/written a letter by the activist. |
19. The deliveryman teased the accountant (who was) saved/given a coupon by her boss. |
20. The prophet spoke of a man (who was) planted/grown a tree by his daughter. |

Appendix 2. Stimuli for Experiments 2 and 3

1. The man reminded us of the defendant/evidence (who/that was) examined the previous day by the lawyer | from the other company.
2. The bandit talked remorsefully of the prisoner/gold (who/that was) transported the whole way by the guards | with horse whips and curry powder.

3. The parents admired the teacher/textbook (who/that was) loved the very best by the class | which came from the ghetto.

4. The foreman worried about the workers/bricks (who/that were) lifted the whole way by the crane | with the faltering engine.

5. The dean asked about the student/paper (who/that was) graded the very lowest by the professor | with a history of discriminating on the basis of nationality.

6. The master of ceremonies discussed the contestant/recipe (who/that was) selected the third time by the judges | who favored originality.

7. The doctor spoke highly of the specialist/equipment (who/that was) requested the second time by the hospital | with the new scanning machine.

8. The police took pictures of the thief/jewelry (who/that was) identified the first night by the victim | from the highrise when a cricket was chirping so loudly in the stairwell.

9. The general thought about the soldier/valley (who/that was) captured the previous day by the troops | with pistols and bayonets.

10. The governor sent reinforcements to the troops/power plant (who were/that was) attacked the previous week by the terrorists | from the higher elevations.

11. The curator criticized the artist/painting (who/that was) studied the previous term by the students | from the university.

12. The governess identified the boy/necklace (who/that was) described the first time by the lady | with the turquoise earrings.

13. Joseph forgot about the mailman/package (who/that was) expected the next day by the secretary | in charge of public relations.

14. The children were worried about the woman/sofa (who/that was) scratched the first day by the cat | with the tawny fur.

15. The agent photographed the man/van (who/that was) recognized the previous day by the spy | from the Russian intelligence bureau.

16. The manager was concerned about the client/account (who/that was) observed the previous month by the agents | from the government.

17. The activist admired the speaker/solution (who/that was) proposed the first time by the group | from the IGC.

18. The congresswoman was referring to the lawyer/package (who/that was) sent the first time by the governor | of Mississippi.

19. The board was concerned about the students/awards (who/that were) accepted the first year by the school | for people with Downs Syndrome.

20. The poet was looking for the woman/portrait (who/that was) sketched the most endearingly by the artist | with the wooden shoes.

Appendix 3

Experiment 1: Means of raw reading times by position (ms).

	Ambiguous		Unambiguous	
Position	Reduced	Unreduced	Reduced	Unreduced
-4	422	438	405	475
-3	462	491	447	476
-2	402	378	390	419
-1	397	440	427	466
0	542	397	475	468
1	632	451	520	471
2	481	455	514	549
3	392	370	393	441
4	367	377	385	397

Experiment 2: Means of raw reading times by position (ms).

	Animate		Inanimate	
Position	Reduced	Unreduced	Reduced	Unreduced
-4	416	426	395	430
-3	434	425	417	428
-2	384	399	357	365
-1	421	405	390	407
0	506	400	449	422
1	553	457	503	443
2	516	480	492	480
3	467	397	433	414
4	414	375	397	396