THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2016
http://dx.doi.org/10.1080/17470218.2016.1186200

Routledge
Taylor & Francis Group

Why reread? Evidence from garden-path and local coherence structures

Kiel Christianson[a,b], Steven G. Luke[c], Erika K. Hussey[b] and Kacey L. Wochna[d]

[a]Departments of Educational Psychology, University of Illinois, Urbana-Champaign, Champaign, IL, USA; [b]Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana-Champaign, Urbana, IL, USA; [c]College of Family, Home, and Social Sciences, Brigham Young University, Provo, UT, USA; [d]Department of Education, Ithaca College, Ithaca, NY, USA

ABSTRACT
Two eye-tracking experiments were conducted to compare the online reading and offline comprehension of main verb/reduced relative garden-path sentences and local coherence sentences. Rereading of early material in garden-path reduced relatives should be revisionary, aimed at reanalysing an earlier misparse; however, rereading of early material in a local coherence reduced relative need only be confirmatory, as the original parse of the earlier portion of these sentences is ultimately correct. Results of online and offline measures showed that local coherence structures elicited signals of reading disruption that arose earlier and lasted longer, and local coherence comprehension was also better than garden path comprehension. Few rereading measures in either sentence type were predicted by structural features of these sentences, nor was rereading related to comprehension accuracy, which was extremely low overall. Results are discussed with respect to selective reanalysis and good-enough processing.

ARTICLE HISTORY
Received 6 July 2015
Accepted 22 April 2016

KEYWORDS
Garden-path sentence; Good-enough processing; Local coherence; Rereading; Language comprehension; Eye movements

Approximately 15% of eye movements during normal silent reading are regressive—that is, moving backwards rather than forwards in the text (Rayner, Pollatsek, Ashby, & Clifton, 2012). The proportion of regressions rises when forward reading is disrupted due to difficulty integrating incoming material. One class of disruptive structures whose effects on eye movements have been extensively studied is temporary structural ambiguities, also referred to as "garden-path sentences" (Bever, 1970). Frazier and Rayner (1982) were the first to systematically investigate eye movements associated with garden paths. The results of that study were interpreted as supporting the selective reanalysis hypothesis, according to which readers regress most often to the portion of the sentence in which the ambiguous choice point was located, rather than just returning to the beginning of the sentence or working stepwise backwards through the text. Most importantly for the present study is that Frazier and Rayner interpreted the eye movement behaviours they observed to indicate reanalysis and correction of the earlier misparse. Frazier and Rayner (1982, p. 182) sum up the selective reanalysis view of regressions and subsequent rereading as follows:

> . . . the parser does not stupidly and automatically proceed through the sentence in one direction or the other regardless of the type of error involved. . . . The selective reanalysis hypothesis thus predicts that the correction procedures . . . may be very specific and will depend on the exact nature of the evidence which the parser has available to it.

In the decades since, numerous researchers have followed Frazier and Rayner's (1982) lead, explicitly linking regressions and rereading to "recovery" from garden paths via structural "reanalysis" (e.g. Ferreira & Clifton, 1986; Ferreira & Henderson, 1991; Frazier & Rayner, 1987; Liversedge, Paterson, & Pickering, 1998; Meseguer, Carreira, & Clifton, 2002; Pickering & Traxler, 1998). Although Frazier and Rayner (1982) acknowledged that certain garden paths were extremely

CONTACT Kiel Christianson kiel@illinois.edu Departments of Educational Psychology, University of Illinois, Education Building, Rm. 240, MC-708, 1310 S. 6th St., Champaign, IL 61820, USA; Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana-Champaign, 405 N Mathews Ave, Urbana, IL 61801, USA

difficult, if not impossible, to recover from (e.g. *The horse raced past the barn fell*; Bever, 1970), the general view is that regressions and rereading are causally linked to recovery from temporary ambiguities, and that recovery is generally achieved. For example, MacDonald, Pearlmutter, and Seidenberg (1994, p. 686) remark 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". This all-or-nothing view of recovery from temporary ambiguities has traditionally been the norm rather than the exception (but see Christianson, Hollingworth, Halliwell, & Ferreira, 2001; and Ferreira, Bailey, & Ferraro, 2002, below).

Although regressing and rereading are not *required* for recovery and accurate comprehension, their occurrence is commonly taken as signalling that readers have noticed the disambiguating error signal and that they are working (somehow) to resolve the poor structural fit between the previous material and the new input. A recent example is the study by Meseguer et al. (2002), who examined rereading patterns as people read mild temporary ambiguities in Spanish. Meseguer et al. observed that approximately 56% of the eye movements from the disambiguating region were regressive. From this region, people tended to regress either to the first verb or to the adverb in the sentence. Either way, these regressions and rereading patterns were associated with "recovery". The authors noted a minimum accuracy rate of 79% correct response to comprehension questions on experimental items (50% chance in all studies cited), but the comprehension questions did not actually probe the attachment site of the ambiguous relative clauses, so there was no way to connect the rereading-related behaviours to the interpretations that they were presumably supporting.

The experiments reported here explicitly test this presumption of recovery, as evidenced by comprehension accuracy, being linked to rereading behaviours and sentence structure. Specifically, we compare rereading in two syntactic structures that share many structural features, but differ with respect to where people should regress to if they are trying to recover from ambiguity. If the targeting of regressions is determined by structural properties of the input, then regression and rereading measures in these structures should differ. Moreover, if the goal of rereading is recovery from initial misinterpretation, then more rereading, especially of the appropriate structurally determined regions, should be associated with better comprehension.

Reduced relative clause temporary ambiguities

In the current work, we compared the rereading and comprehension of sentences such as (1–2). In both, there is a temporary ambiguity of the same sort: a reduced relative (RR) clause. In (1), the RR is in subject position, which results in a garden-path sentence (henceforth GP RR) of the sort examined by Frazier and Rayner (1982, 1987) and many others in the years since. These sentences are parsed on the first pass with the relative clause verb (*knitted*) as the main verb. When the actual main verb (*waved*) is encountered, however, the initial analysis becomes unlicensed. In order to incorporate the new main verb into the structure, the reader must recognize that there was a reduced relative clause (*The woman* [*who was*] *knitted . . .*). Depending on one's preferred view of the parser, there are a number of competing theories about how the recovery sketched above might take place (e.g. Abney, 1989; Altmann & Steedman, 1988; Frazier, 1998; Frazier & Clifton, 1996; Frazier & Fodor, 1978; Frazier & Rayner, 1982, 1987; Gorrell, 1998; MacDonald et al., 1994; Pritchett, 1992; Spivey-Knowlton & Sedivy, 1995; Townsend & Bever, 2001). In the current work, we remain agnostic as to a specific parsing theory. The important point is that all of them predict that sentences such as (1) should cause disruption to processing, and this disruption leads to increased rates of regressions and rereading, compared to non-garden-path sentences.

(1) The woman knitted the scarf for the ceremony waved at the mayor.
(2) The mayor waved at the woman knitted a scarf for the ceremony.

Sentence (2) represents another type of disruptive sentence, a local coherence (LC) structure (Tabor, Galantucci, & Richardson, 2004). LCs also contain RR clauses (henceforth LC RR), but they occur in object position, rather than subject position. This structural difference is critical, as the RR in object position should not be disruptive to processing, as it is the only licensed structural option for the continuation of the sentence at this point in the parse. In other words, because the main verb has already been encountered, the only licit interpretation for *knitted* in (2) is as the verb of a RR clause. Despite this structural constraint, Tabor et al. (2004) demonstrated in a series of self-paced reading experiments that processing was disrupted at and immediately after the LC RR verb. Tabor et al. argued that current parsing theories have difficulty predicting this disruption and proposed various flavours of self-organized parsing accounts, to which we return in the General Discussion. For now, we limit our consideration of these structures to the predictions about rereading that we derive directly from their structures. If rereading-related eye movement measures index recovery processes, these measures should differ in certain ways between GP RRs like (1) and LC RRs like (2). Moreover, if recovery is successful, readers should be able to answer comprehension questions relatively accurately about the material they have read and reread in both structures, and eye movements corresponding to rereading should be related to comprehension accuracy. The goal of the work reported here was to test both of these hypotheses.

One aspect of LC RRs that is not clear is how much, if at all, the processing of LC RRs differs from that of GP RRs. As noted above, Tabor and colleagues (2004) observed that the LC structure is interesting because the earlier content of the sentence should rule out or block the misparse later in the sentence, but it appears not to. Yet there is another way of characterizing LC RRs when they are directly compared to GP RRs. Both structures provide the opportunity for local coherence, as misparses in both are caused by the linear adjacency of lexical items that are not syntactically adjacent. The important distinction is the placement of the locally coherent string. In traditional GP RRs, the misleadingly locally coherent (ambiguous) string is located earlier in the sentence, in subject position, and the disambiguating information follows the main verb. GP RRs, therefore, require reanalysis (or reranking) of *earlier* structure in order to create (or promote) a structure that is fully licensed by the grammar and an interpretation that is faithful to the actual content of the input. LC RRs, on the other hand, contain the misleadingly locally coherent (termed here as "ambiguous" to be consistent with the GP RR terminology) string later in the sentence, in object position, preceded by unambiguous syntax, and the main verb. In GP RRs, a reader has no way of knowing that his or her parse is incorrect until the main verb is encountered; the error comes first and must be repaired if the correct interpretation is to be derived. In LC RRs, readers begin with the correct structure, but it is one that necessitates a specific parse of the ending that is infrequent or noncanonical.

The distinction drawn here between the two structures—that LC RRs do not require revision of early parse assumptions, whereas GP RRs do—is important for the following reason. In the GP RRs, the disambiguating word [*waved at* in (1)] is an error signal that tells the parser that the current structure is no longer tenable. Rereading is therefore generally taken as evidence of reanalysis of the initial incorrect parse so that the ongoing information (in this case, the main verb of the sentence) can be integrated into a coherent global structure. As noted, researchers operating under this set of assumptions about rereading often talk about recovery from garden paths (e.g. Ferreira & Henderson, 1991; Liversedge et al., 1998; Pickering & Traxler, 1998) and discuss rereading-related eye movements explicitly in terms of reanalysis (e.g. Frazier & Rayner, 1982; Liversedge et al., 1998; Meseguer et al., 2002).

In contrast, in the LC RRs the error signal is the relative clause verb [*knitted* in (2)]. Recall that this error signal causes disruption, even though it should not (Tabor et al., 2004) given the unambiguous nature of the preceding sentence context (i.e., there is already a main verb, so *knitted* must be the verb in a RR clause). Rereading prior to *the woman* in LC RRs, then, cannot signal reanalysis in the classic sense of restructuring. Instead, rereading in LC RRs might be signal checking, or confirming, the structure and content of what has been previously read. Thus, the same qualitative behaviour(s)—eye movements associated with rereading—should differ between these sentence types, given that, theoretically, very different processes are being carried out in the two sentence types.

Note that the above characterization is consistent with proposals by Gibson (1998, 2000) about why both reduced and unreduced relative clauses that modify subjects should be harder to process than those that modify objects. Under Gibson's dependency locality theory, which operationalizes processing difficulty in terms of memory and integration costs, both ambiguous and unambiguous RRs in subject position impose greater processing memory cost than do relative clauses in object position. The reason for this is that more referents intervene between the subject and the main verb in subject-modifying relative clauses than in object-modifying relative clauses. Grodner, Gibson, and Tunstall (2002) presented reduced and unreduced relative clauses modifying both subjects and objects in a self-paced reading experiment. They observed that ambiguous RRs led to longer reading times (taken as an index of reading difficulty) than unambiguous relative clauses, and also that subject-modifying relatives led to longer reading times than object-modifying relatives. Furthermore, they noted that ambiguity effects were larger for subject-relatives than object-relatives. Of particular relevance to the present study, Grodner et al. argued that a critical aspect of reanalysis is "how committed the perceiver is to the target and alternative structures", and furthermore that they knew of "no theory of reanalysis that incorporates this type of information" (p. 288). However, at least since Ferreira and Henderson (1991), it has been observed that the longer the parser is committed to an incorrect analysis, the harder it is to recover from it (cf. Christianson et al., 2001; Tabor & Hutchins, 2004). Indeed, Kimball (1973) proposed that there is a cost associated with recalling a previous phrase out of memory to reanalyse its structure. As such, it is reasonable to expect that GP RRs should be more difficult to recover from than LC RRs.

In any case, despite the important structural difference between GP RRs and LC RRs, there is, to our knowledge, no direct evidence that the two types of processing difficulties result in different patterns of rereading or comprehension accuracy. Moreover, despite some discussions in the literature cited above about reanalysis and recovery from both GP RRs and LC RRs, these processes are often not examined (e.g. in self-paced reading paradigms in which rereading is not possible), and "recovery" is ill-defined (e.g. not measured by probing final interpretations). This gap in the literature is theoretically important, as the rereading behaviours generally observed when people read these sentences are assumed to reflect underlying recovery processes: Substantial theoretical claims have been made in the syntactic parsing literature based on the observation that people usually interrupt their forward reading through these sentences and reread some or all of the previous input (when allowed by the paradigm).

Meseguer et al. (2002), discussed above, use the term "overt reanalysis strategy" to describe rereading that consists of regression back to the original choice point of relatively mild garden paths in Spanish and the subsequent rereading of the input thereafter. They use the term "covert reanalysis strategy" to describe perseverance on the error signal while alternative parses are retrieved from memory without going back to reread. Meseguer et al. also explicitly "expected the eye movement behavior observed . . . to inform us about how reanalysis processes are successfully applied, not how they go astray" (Meseguer et al., 2002, p. 553); however, there was no direct evidence of successful reanalysis. Nevertheless, because of where the regressions were (mostly) targeted, Meseguer et al. argued that their results supported the selective reanalysis hypothesis (Frazier & Rayner, 1982).

There is evidence that rereading improves comprehension of and memory for texts (e.g. Amlund, Kardash, & Kulhavy, 1986; Calvo, 2001; Schotter, Tran, & Rayner, 2014). For example, Schotter et al. (2014) used a trailing mask paradigm to show that comprehension suffers when readers are denied the option of rereading. However, their results failed to show any difference in comprehension between trials in which rereading was undertaken and trials in which it was not when previous text had not been masked. In other words, their results suggest that readers need not necessarily reread, but when they do, it is because they have failed to comprehend or process adequately on the first pass, and rereading is used as a recovery strategy.

Von der Malsburg and Vasishth (2011, 2013) report a novel method of analysing regression and rereading patterns called scanpath analysis. In their 2011 paper, von der Malsburg and Vasishth showed considerable variation in rereading, but with three prototypical scanpath patterns in a reanalysis of the Meseguer et al. (2002) data: regression back to the beginning of the sentence followed by rereading of the entire sentence, a single saccade to an early region in the sentence, and a regression to the nearby disambiguating region. In their 2013 paper reporting the results of an original experiment on Spanish adjunct attachment, the authors again observed three prototypical scanpath patterns comparable to the 2011 results. They furthermore associated these with first-pass reading patterns and showed that readers who slowed down at potential attachment site during the first pass were more likely to regress and reread. Consistent with good-enough processing theory (Christianson, 2016; Christianson et al., 2001; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Patson, 2007; Swets, Desmet, Clifton, & Ferreira, 2008), von der Malsburg and Vasishth found that readers with more working memory capacity were more likely to attach and, therefore, more likely to regress when that attachment was found to be incorrect. Lower capacity readers, who read faster on first pass and by assumption did not attach, were less disturbed by later material that would not have been consistent with a previous (high) attachment.

Importantly, von der Malsburg and Vasishth (2013) also probed the comprehension of the attachment and found 58% accuracy in high-attachment conditions (where an initial attachment was ultimately incorrect), and 70% in low-attachment conditions (where an initial attachment was ultimately correct). Relatively low rates of comprehension accuracy are expected under good-enough (and some other) parsing accounts (see below). They did not analyse the data in such a way as to be able to draw any conclusions about the connection between scanpath pattern and accuracy, however. What the low (but above chance) accuracy in the most difficult condition did show was that misinterpretations seem to persist even when readers regress and reread. Unlike the adjunct attachment structures examined by von der Malsburg and Vasishth, the RR structures examined here are less likely to be structurally underspecified on first pass, as the ambiguities involve thematic assignments and verb–argument relations (e.g. Frazier & Clifton, 1996), and thus regressions should be even more common.

Persistent misinterpretation effects

The reasoning that regressions and rereading can be interpreted as indices of "recovery" holds as long as we assume that readers do indeed (usually) eventually arrive at an interpretation of both GP RRs and LC RRs that are faithful to the licensed structure of the input. This assumption has recently has come into question. Tabor et al. (2004) collected grammatical acceptability ratings for their LC structures and found them to be very low across the board, with some acceptability rates as low as 22% on in binary yes/no forced choice (though acceptability and comprehension are two very different things). Numerous recent studies strongly suggest that readers do not consistently derive the structurally licensed interpretations from garden-path sentences such as (3a) (e.g. Christianson et al., 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira et al., 2001; Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013), or other challenging but unambiguous sentences (Christianson, Luke, & Ferreira, 2010; Ferreira, 2003; Zhou & Christianson, 2016).

(3) a. While Mary dressed the baby played in the crib.
b. Did Mary dress the baby?
c. Did the baby play in the crib?

It has been proposed that reanalysis is an operation of "last resort" (Fodor & Frazier, 1980; Sturt, Pickering, Scheepers, & Crocker, 2001), as the parser appears loath to relinquish a plausible misparse (Ferreira & Henderson, 1991). Christianson et al. (2001) first pointed out that the final mental representations of garden paths are unfaithful to the actual content of the input more often than previously suspected. The example in (3) is one of the garden-path sentences used by Christianson et al. (2001; see also Christianson et al., 2006; Hussey, Ward, Christianson, & Kramer, 2015; Slattery et al., 2013). The authors found that people are likely to answer "Yes" to both comprehension questions like (3b) and (3c) after reading sentences like (3a), and are equally confident of both answers. These results were taken as being suggestive of incomplete syntactic reanalysis, resulting in a final representation that allowed *the baby* to be both object of the subordinate verb *dressed* and the matrix verb *played*—an interpretation that is unlicensed by the syntax. In other words, it appeared as if people were deriving an interpretation that was consistent with two overlapping substrings within sentences like (3a), but that one integrated, global interpretation was not being derived. It has subsequently been shown that similar misinterpretation effects can be observed in paradigms that do not involve explicit comprehension probes—for example, structural priming (van Gompel, Pickering, Pearson, & Jacob, 2006), paraphrasing (Patson, Darowski, Moon, & Ferreira, 2009), and eye tracking (Slattery et al., 2013).

Such misinterpretation effects are problematic for many current models of human sentence processing. Irrespective of theoretical persuasion, most current models assume, at least implicitly, that aside from scattered anomalies, sentence structure is successfully and completely computed, and an interpretation faithful to the actual sentence content is generated from that structure. If this does not occur, the parse, and comprehension, are widely assumed to fail. Exceptions to this all-or-nothing assumption (in addition to good enough; cf. Traxler, 2014) include models that postulate some sort of uncertainty in the mind of the reader about previously encountered material (e.g. Gibson, Bergen, & Piantadosi, 2013; Levy, 2008; Levy, Bicknell, Slattery, & Rayner, 2009). These so-called "noisy-channel" models are predicated on the fact that language is fast, speakers and texts vary, and comprehenders have limited memory. As such, when faced with incongruent incoming material, comprehenders might misremember previous input, and "edit" their memory representations to fit previous material with incoming material. We return to these models in the General Discussion.

In summary, rereading-related eye movements have been taken as strong evidence that readers notice disruptions and recover from them successfully, when it is also a possibility that rereading indexes efforts to recover from misanalysis regardless of their ultimate success. Furthermore, rereading might simply be done to confirm the reader's memory for what had been previously read, irrespective of whether the structure or interpretation of the previous material was actually correct within the context of the full sentence.

Current study predictions

How then can comparing reading and comprehension of GP RRs to LC RRs inform our interpretation of rereading-related eye movements? If rereading in temporarily ambiguous sentences is driven by strategic reanalysis aimed at revising the initial parse (e.g. Frazier & Rayner, 1982, 1987; Meseguer et al., 2002; cf. Mitchell, Shen, Green, & Hodgson, 2008), then we should see regressions into Region 2 [*tossed* in (4a)] and subsequent rereading from that point onward only in GP RRs, as they require reanalysis. Note that we make predictions here only about temporarily ambiguous structures, and "selective reanalysis" does not generate any predictions about structures that do not need to be reanalysed. (We might also see regressions in GP RRs back to Region 1, if regressions are mistargeted, or if a complete rereading is required to revise the structure; cf. von der Malsburg & Vasishth, 2013.) LC RRs, on the other hand, should show fewer long-range regressions and shorter rereading time on previous parts of the sentence, as they should not require structural reanalysis on the view that only one parsing option (RR) is available to the parser (Tabor et al., 2004). Instead, LC RRs should trigger perseverance at the error signal [Region 4, *tossed* in (4c)], which would signal that the parser is figuring out how to integrate apparently ill-fitting material into the previously correct structure. We might also observe short regressions into the immediately preceding noun phrase [*the player* in (4c)], analogous to the way the Region 1 noun phrase might be targeted in RR GPs: as a sort of springboard into the critical choice point. Alternatively, a regression to this noun phrase could be interpreted as a simple "time-out" while the parser catches up to the eye (Mitchell et al., 2008).

It might be, though, that regressing and rereading are not done exclusively to reanalyse, but rather to confirm, or check, that previously read material is in fact what had been previously read (cf. Levy et al., 2009). Under this more confirmatory view, rereading in LC RRs will be relatively more successful, but rereading in GP RRs—in which a misparse will be confirmed—will result in misinterpretation. In other words, if readers are unsure that they have read the string *the player tossed the ball* in (4a, below), rereading this string will confirm that that was indeed the input. We assume also that, if the identity of the string is confirmed, its original (likely incorrect) analysis will also be confirmed, and thus reinforced, leading by hypothesis to mostly incorrect interpretations. On the other hand, if readers reread the previous content of the RR LC (4c, below)—*The other team interfered with the player*—to confirm the contents, the contents will be accurately confirmed. Moreover, doing so will reinforce the ultimately correct parse of that string, allowing the attempt to integrate the incoming LC material—an attempt that should be more successful than in GP structures, but still might be relatively error-prone, given the complexity and infrequency of these structures.

There is no current eye movement theory that describes qualitative or quantitative differences between rereading measures driven by attempts to confirm the previous parse/interpretation versus those driven by attempts to reanalyse the previous parse/interpretation. However, if comprehension accuracy rates to questions such as *Did the woman knit the scarf?* [for (1) and (2) above] are lower for GP RRs than for LC RRs, we could infer that confirmatory rereading is generally more successful than revisionary rereading. Accuracy rates for GP RRs and LC RRs around chance levels would suggest frequent parsing failure—either of revision or of integration, respectively—and guessing borne out of confusion (perhaps accompanied by inflated question response latencies). Accuracy rates below chance, however, would suggest perseverant misinterpretation, not just confusion. In the GP RR condition in particular, accuracy below chance could be taken as evidence of misapplied confirmatory rereading. If reanalysis is a "last resort" (Sturt et al., 2001), then there may in fact be a general preference for confirmatory rereading, irrespective of what the structure calls for. Ferreira and colleagues predict just such a preference within the good-enough theory (Christianson et al., 2001; Christianson et al., 2010; Christianson et al., 2006;

Ferreira et al., 2002; Ferreira & Patson, 2007), as long as the confirmatory process returns a parse that is relatively plausible, even if it is not globally coherent. Finally, given that the parser appears loath to relinquish early interpretations, as well as structures, comprehension of LC RRs might be more accurate than that of GP RRs. A comprehension advantage for LC RRs is also predicted by Gibson's dependency locality theory (1998, 2000; Grodner et al., 2002).

In summary, one goal of the present study was to compare rereading behaviours in two types of sentences that generally trigger rereading-related eye movements. Importantly, the structures differ with respect to what processes need to be assumed to operate during rereading if successful interpretation is to be achieved. In GP RR sentences, rereading should be revisionary. In LC RR sentences, rereading, if it occurs at all, should be confirmatory. Another goal was to determine whether there is any reliable connection between how these sentences are read online and how accurately people respond to comprehension questions about them offline. As far as we know, there has been no systematic comparison of the online rereading or offline comprehension of RR garden paths and LC structures, nor have comprehension rates of sentences such as (1) and (2) been compared, or usually even collected (cf. Tabor et al.'s, 2004, acceptability data). This would seem to be a serious oversight, given that both sentence types are used to derive evidence in favour of (or against) so many theories of human sentence processing.

Experiment 1

Method

Participants

Forty-eight people from the University of Illinois at Urbana-Champaign community participated. The large majority of participants were recruited from the Educational Psychology subject pool. All participants were compensated for their time with either course credit or $7.

Apparatus

Eye-movements were recorded via an SR Research Ltd. Eyelink 1000 eye tracker, which records the position of the reader's eye once every millisecond (1000 Hz sampling rate), and has a high spatial resolution of 0.01°. Text was displayed in 12-point Courier New font. Participants were seated 70 cm away from a 20″ monitor. At this distance, approximately 3.5 characters subtended 1° of visual angle. Head movements were minimized with chin and head rests. Although viewing was binocular, eye movements were recorded from the right eye.

Materials

Forty items were created, half of which were garden-path sentences, and half of which were local coherence sentences, which were equated pairwise for lexical content. A full example set can be found in (4). Note that pipes (|) delineate regions of interest that are referred to in subsequent analyses. Counting from left to right, ambiguous sentences (4a, 4c) have 5 regions, and unambiguous sentences (4b, 4d) have 6. The critical region in ambiguous items is Region 4; in unambiguous items it is Region 5. Pre-critical regions in ambiguous items are Regions 1–3. The spillover region in all items is the final region.

(4) (a) The player | tossed | the ball | *interfered with* | the other team.
(b) The player | who was | tossed | the ball | *interfered with* | the other team.
(c) The other team | interfered with | the player | *tossed* | the ball.
(d) The other team | interfered with | the player | who was | *tossed* | the ball.
(e) Did the player toss the ball?

This resulted in a 2 (sentence type: garden path vs. local coherence) × 2 (ambiguity: ambiguous vs. unambiguous) design. The 40 items were counterbalanced across four list conditions in a Latin square design.

Procedure

Each trial involved the following sequence: Each trial began with a gaze trigger, which consisted of a black circle 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 a standard game controller to indicate that she or he had finished reading the sentence. At this point, the sentence disappeared. Next, a question about the content of the sentence appeared, which participants answered by pressing the appropriate button on the controller. Then the next trial began. Sentences were presented in a random order for each participant, and each testing session lasted approximately 40 minutes. In addition to the 40 experimental items, each list contained 72 other sentences with a variety of structures, all of which were followed by comprehension probes of some sort. Eighty percent of these other sentences were items from unrelated experiments, including more or less plausible subject- and object-relative clauses, sentences containing direct and indirect quotes, and direct object/ sentence garden paths conjunction. Twenty percent were true fillers, consisting of unambiguous complex sentences. The correct answer to 50% of the items across all item types was Yes; that to the other 50% was No. Mean accuracy on true filler sentences in Experiment 1 was 84%; mean accuracy for all items excluding the items for this experiment was 84% (range: 52–91%).

Results

Online data

Because the critical regions in the garden-path and local coherence conditions differed in length, any comparison based on untransformed reading time measures would have been difficult to interpret. For this reason, residual reading times were used as the dependent measure, as advocated by Trueswell, Tanenhaus, and Garnsey (1994). Residual reading times were calculated by fitting a separate linear mixed model for each reading time measure, with length (in characters) of the region as the fixed effect and with random intercepts and slopes for each participant. For go-past time, the length of the entire preceding sentence was used instead, as is standard for computing go-past time, which is the sum of all fixations up to and including on the target region until the eyes have moved off it to the right, including refixations on previous material (Rayner, 1998). Go-past times were also analysed in two ways. The first used residualized reading times because go-past times are likely to correlate with the length of the preceding material. Using residual reading times permits a more direct comparison of the two sentence types and should reveal any differences in processing difficulty between the two sentences. The second used raw reading times because the first-pass component of go-past times is unlikely to be affected by the length of the preceding material. (As seen in the analyses that follow, these two analyses returned nearly identical results.)

In the GP conditions, the critical region was the disambiguating verb [*interfered with* in (4a) and (4b)]. In the LC conditions, this region was the relative clause verb [*tossed* in (4c) and (4d)], which was locally coherent with the direct object of the main verb (*player*) in the ambiguous condition. For both types of sentences, it was expected that participants would experience processing difficulty at the critical region, and this difficulty should manifest itself in inflated reading times in the ambiguous condition. Descriptive statistics for raw reading times are reported in Tables 1 and 2.

For the analysis of the online data, five different dependent measures of reading time were analysed in the critical and spillover regions using repeated measures analyses of variance (ANOVAs): first-fixation duration, gaze duration, go-past time, total time, and first-pass regression out probability. We also examined rereading time and the probability of regressing into each pre-critical region. For each ANOVA, list condition was included as a between-participants and between-items factor (Pollatsek & Well, 1995) to remove variance associated with the individual lists. The effects from all ANOVAs can be found in Table 3 in the left panels. Before analysis, fixations shorter than 80 ms and longer than 1200 ms were excluded from the data set (4% of fixations). One faulty item was excluded from the analysis due to a typo in a critical word, leaving 39 items in total.

Critical region

First-fixation duration. For first-fixation duration (the length of the first fixation on the region), the main effect of sentence type was not significant. However, there was a main effect of ambiguity, with longer fixations in ambiguous sentences, as well as a significant interaction of ambiguity and sentence type. Follow-up *t* tests indicated that the effect of ambiguity was only significant for the LC sentences [$t_1(47) = 3.74$, $p = .001$; $t_2(38) = 5.82$, $p < .0001$], and not for the GP sentences (both $ts < 0.82$).

Gaze duration. For gaze duration (the sum of all fixations on a region before leaving it in either direction), there was a main effect of sentence type by participants only, such that GPs had longer gaze durations than LCs. The main effect of ambiguity was also significant, indicating longer reading times for ambiguous sentences. The interaction was significant by participants but not by items. This interaction suggests that the effect of ambiguity was strong in the LC condition [$t_1(47) = 5.59$, $p < .0001$; $t_2(38) = 4.84$, $p < .0001$], but only marginal in the GP condition [$t_1(47) = 1.7$, $p = .092$; $t_2(38) = 1.74$, $p = .087$].

Go-past time. For go-past time (the sum of all fixations made after the eyes first enter a region from

Table 1. Mean performance for eye movement data in the critical region for Experiments 1 and 2: Raw reading times.

Type	Ambiguity	First-fixation duration (ms)		Gaze duration (ms)		Go-past time (ms)		Total time (ms)		First-pass regression out probability	
		Exp1	Exp2	Exp1	Exp2	Exp1	Exp2	Exp1	Exp2	Exp1	Exp2
Critical region											
GP	Ambiguous	258 (117)	222 (34)	391 (259)	480 (147)	773 (1149)	798 (361)	1126 (835)	1252 (488)	.28 (.16)	.26 (.20)
	Unambiguous	253 (110)	233 (41)	357 (213)	431 (129)	599 (647)	629 (234)	790 (524)	866 (263)	.24 (.15)	.17 (.15)
LC	Ambiguous	275 (121)	257 (48)	392 (277)	327 (74)	1096 (1686)	765 (406)	1092 (785)	868 (395)	.35 (.19)	.33 (.17)
	Unambiguous	235 (93)	233 (37)	298 (177)	277 (61)	588 (767)	577 (280)	711 (474)	608 (191)	.27 (.19)	.38 (.22)
Spillover region											
GP	Ambiguous	235 (47)	225 (40)	375 (170)	326 (109)	2722 (1591)	2233 (1242)	853 (349)	691 (268)	.87 (.16)	.87 (.13)
	Unambiguous	250 (46)	239 (51)	437 (162)	379 (149)	2108 (1034)	1519 (741)	712 (287)	584 (273)	.81 (.21)	.80 (.20)
LC	Ambiguous	243 (48)	236 (49)	373 (133)	362 (147)	2950 (1681)	2487 (1492)	831 (382)	755 (310)	.88 (.19)	.85 (.16)
	Unambiguous	243 (63)	228 (41)	412 (205)	331 (97)	2503 (1353)	1902 (850)	777 (334)	606 (237)	.88 (.17)	.87 (.14)

Note: GP = garden path; LC = local coherence; Exp = Experiment. Standard deviations in parentheses.

the left but before leaving a region to the right, including any refixations on earlier regions resulting from regressions out of the region; also called regression-path duration), the main effect of sentence type was significant, indicating that more time was spent rereading after first encountering the relative clause verb in the LCs. The main effect of ambiguity was also significant, indicating longer fixations for ambiguous sentences. The interaction was significant by both participants and items. Follow-up two-tailed *t* tests indicated that the effect of ambiguity was larger for LC sentences [$t_1(47) = 4.14$, $p < .0001$; $t_2(38) = 4.04$, $p < .0001$] than for GP sentences [$t_1(47) = 1.90$, $p = .06$; $t_2(38) = 2.58$, $p = .041$].

The preceding analyses were computed over go-past times that were residualized over the length of the sentence up to and including the critical region. A second set of analyses was performed using raw (unresidualized) go-past times. These analyses replicated the analyses of residualized times: The effect of sentence type was only present for ambiguous sentences [$t_1(48) = -2.62$, $p = .012$; $t_2(39) = -1.63$, $p = .11$], and not for unambiguous sentences [$t_1(48) = -0.31$, $p = .76$; $t_2(38) = 0.13$, $p = .90$]. The effect of ambiguity was larger for LC sentences [$t_1(48) = 5.36$, $p < .0001$; $t_2(39) = 3.43$, $p = .001$] than for GP sentences [$t_1(48) = 3.36$, $p = .002$; $t_2(38) = 3.66$, $p = .0008$].

Total time. For total time (the sum of all fixations on a region), the main effect of sentence type was only marginally significant by participants and non-significant by items. The main effect of ambiguity was highly significant, indicating longer fixations for ambiguous sentences. The interaction was not significant.

Probability of first-pass regressions out of the critical region. Analyses of the probability that a regression would be launched from the critical region to a previous part of the sentence (prior to moving forward from that region) revealed a main effect of sentence type by subjects but not by items, indicating that more first-pass regressions were made from the critical region of LCs. There was also a main effect of ambiguity, indicating that more first-pass regressions were made from ambiguous critical regions. There were no significant interactions.

Pre-critical regions

For analyses of pre-critical regions, we initially focused our attention on ambiguous items, given that we are

Table 2. Mean performance for eye movement data in the critical region for Experiments 1 and 2: Residualized reading times.

		First-fixation duration (ms)		Gaze duration (ms)		Total time (ms)	
Type	Ambiguity	Exp1	Exp2	Exp1	Exp2	Exp1	Exp2
Critical region							
GP	Ambiguous	27.29 (41.8)	4.97 (26.14)	28.17 (83.16)	93.34 (101.04)	282.42 (245.5)	445.82 (268.52)
	Unambiguous	23.15 (30.52)	14.89 (31.78)	−1.52 (76.45)	44.6 (78.97)	−45.93 (247.32)	61.6 (179.26)
LC	Ambiguous	45.07 (39.34)	37.39 (36.53)	71.85 (80.36)	57.8 (50.71)	314.57 (233.8)	286.67 (220.38)
	Unambiguous	4.34 (34.01)	13.71 (28.98)	−7.15 (63.34)	3.07 (55.98)	−39.16 (188.39)	17.27 (126.63)
Spillover region							
GP	Ambiguous	10.58 (40.61)	7.34 (29.13)	−96.75 (133.65)	−64.17 (73.83)	−166.83 (256.04)	−121.98 (211.47)
	Unambiguous	24.86 (36.38)	21.38 (42.58)	−31.31 (115.55)	−11.65 (102.45)	−305.41 (296.6)	−228.13 (247.81)
LC	Ambiguous	17.15 (39.03)	18.41 (36.56)	−79.73 (124.16)	−15.87 (104.32)	−138.84 (310.29)	−25.75 (240.18)
	Unambiguous	17.23 (49.72)	10.44 (36.9)	−42.65 (140)	−46.67 (82.56)	−208.83 (266.83)	−173.15 (219.51)

Note: GP = garden path; LC = local coherence; Exp = Experiment. Standard deviations in parentheses.

Table 3. ANOVA effects at the subject level and item level for each eye movement measure for Experiments 1 and 2.

	Experiment 1				Experiment 2			
Measure	$F_1(1, 44)$	p_1	$F_2(1, 35)$	p_2	$F_1(1, 44)$	p_1	$F_2(1, 36)$	p_2
	Critical region							
First-fixation duration								
Sentence type	0.01	.905	0.00	.978	**5.40**	**.025***	**4.27**	**.046***
Ambiguity	**14.36**	**.000***	**24.40**	**.000***	1.88	.177	2.79	.104
Sentence Type × Ambiguity	**13.39**	**.001***	**9.52**	**.002***	**19.04**	**.000***	**16.35**	**.000***
Gaze duration								
Sentence type	**4.45**	**.041***	2.67	.111	**8.53**	**.006***	**5.54**	**.024***
Ambiguity	**36.22**	**.000***	**31.04**	**.000***	**21.10**	**.000***	**62.97**	**.001***
Sentence Type × Ambiguity	**6.92**	**.012***	2.60	.120	0.47	.830	0.79	.840
Go-past time								
Sentence type	**32.00**	**.000***	**11.62**	**.002***	0.01	.925	0.06	.808
Ambiguity	**23.80**	**.000***	**31.47**	**.000***	**7.37**	**.009***	**9.96**	**.003***
Sentence Type × Ambiguity	**9.31**	**.004***	**4.20**	**.047***	0.07	.799	0.00	.995
Total time								
Sentence type	3.440	.070†	0.020	.890	3.21	.080†	2.36	.134
Ambiguity	**52.52**	**.000***	**129.55**	**.000***	**62.58**	**.000***	**127.24**	**.000***
Sentence Type × Ambiguity	0.32	.577	0.03	.860	**8.25**	**.006***	**7.46**	**.010***
First-pass regressions out								
Sentence type	**9.69**	**.003***	2.85	.100	**53.81**	**.000***	**19.92**	**.000***
Ambiguity	**6.84**	**.012***	**5.26**	**.028***	0.33	.567	0.06	.817
Sentence Type × Ambiguity	1.33	.255	0.93	.343	**9.45**	**.004***	**13.54**	**.001***
	Spillover region							
First-fixation duration								
Sentence type	0.01	.933	0.01	.945	0.00	.987	0.01	.930
Ambiguity	1.84	.181	1.82	.186	0.28	.597	0.32	.577
Sentence Type × Ambiguity	2.02	.163	**5.95**	**.020***	**4.94**	**.031***	**6.54**	**.015***
Gaze duration								
Sentence type	0.05	.820	0.00	.973	0.42	.519	0.36	.553
Ambiguity	**13.43**	**.001***	**12.36**	**.001***	1.02	.319	0.97	.331
Sentence Type × Ambiguity	2.28	.138	1.49	.231	**12.58**	**.001***	**14.05**	**.001***
Go-past time								
Sentence type	**23.42**	**.000***	**11.87**	**.002***	**25.68**	**.000***	**5.48**	**.025***
Ambiguity	**14.52**	**.000***	**76.60**	**.000***	**29.92**	**.000***	**63.95**	**.000***
Sentence Type × Ambiguity	1.00	.322	1.05	.313	1.06	.309	1.97	.170
Total time								
Sentence type	**8.40**	**.006***	1.71	.200	1.51	.226	0.54	.470
Ambiguity	**16.72**	**.000***	**28.42**	**.000***	1.31	.258	0.96	.335
Sentence Type × Ambiguity	2.53	.119	1.13	.295	**8.06**	**.007***	**6.50**	**.016***
First-pass regressions out								
Sentence type	3.86	.056†	**7.99**	**.008***	**20.20**	**.000***	**15.14**	**.000***
Ambiguity	2.97	.092†	**5.01**	**.032***	**26.49**	**.000***	**47.56**	**.000***
Sentence Type × Ambiguity	2.93	.094†	**5.98**	**.020***	0.71	.405	0.36	.552

Note: Subject level: F_1, p_1; item level: F_2, p_2. Bold indicates reliable effects. ANOVA = analysis of variance.
$^*p < .05$. $^{\dagger}p < .10$.

concerned here with rereading behaviours and the connection between rereading, "recovery" from temporary ambiguity, and comprehension accuracy. Furthermore, the "selective reanalysis hypothesis" (Frazier & Rayner, 1982), which largely motivates the current investigation, makes specific predictions about reanalysis from temporary ambiguities, but none about rereading of unambiguous structures. Analyses of first-pass measures on critical and spillover regions yielded robust and consistent ambiguity effects. As such, if participants did reread in unambiguous items, it was, by hypothesis, for reasons other than for recovering from a temporary ambiguity. Nevertheless, we also compared unambiguous controls and ambiguous items on all measures to determine whether ambiguity interacted with structure and to ensure that item-specific differences (e.g. plausibility) did not drive results in pre-critical regions.

Probability of regressing into pre-critical regions. The probability of regressions into pre-critical regions was analysed for each pre-critical region in ambiguous items only. Recall that the selective reanalysis hypothesis (Frazier & Rayner, 1982) predicts that readers should direct regressions differentially in GP RRs and LC RRs. Specifically, in ambiguous GPs, Region 1 (*The player*) and Region 2 (*tossed*) should be targeted for regressions. If there are any reanalysis-related regressions at all in ambiguous LCs, Region 3 (*the player*) should be targeted as the beginning of the RR. The left panel of Figure 1 illustrates the mean probabilities for regressions into these (shaded) pre-critical regions.

We observed a main effect of sentence type in Region 1, indicating a greater probability of regressing into the first region of LC sentences [$F_1(1, 45) = 12.08$, $MSE = .206$, $p = .001$; $F_2(1, 36) = 14.39$, $MSE = .185$, $p < .0001$]. This same pattern was observed when analysing unambiguous controls [$F_1(1, 45) = 6.15$, $MSE = .100$, $p = .017$; $F_2(1, 36) = 4.99$, $MSE = .094$, $p = .033$], suggesting that participants tended to reread LCs from the beginning more often whether they were ambiguous or not. We found no effect of sentence type in any other pre-critical region of ambiguous sentences (F_1s $< .900$, $ps > .34$; F_2s $< .685$, $ps > .41$); however, in unambiguous sentences, more regressions into the verb region were observed in LC than in GP sentences [$F_1(1, 45) = 10.12$, $MSE = .228$, $p = .003$; $F_2(1, 36) = 7.75$, $MSE = .187$, $p = .008$]. Accordingly, there was a marginally significant interaction between sentence type and ambiguity [$F_1(1, 45) = 3.97$, $MSE = .069$, $p = .052$; $F_2(1, 36) = 3.91$, $MSE = .066$, $p = .056$] such that the ambiguity effect (ambiguous minus unambiguous regressions-in) for GP sentences (effect = .126) was larger than that for LC sentences (effect = .051). Given that accuracy rates in question response across sentence types in the unambiguous condition was not significant, it appears that differences in regressions into the verb region in unambiguous items did not influence comprehension. Similarly, the slightly higher count of regressions into the verb region in ambiguous GP sentences did not improve comprehension for these structures. No other significant differences in regressions-in were found in any pre-critical regions of the unambiguous controls (e.g. relativizer or second noun phrase regions; F_1s < 2.7, $ps > .10$; F_2s $< .800$, $ps > .37$).

Rereading times in pre-critical regions. We also evaluated the time that subjects spent rereading (sum of all fixations on a region subsequent to first pass) regions prior to the critical region (left panel of Figure 2). Analyses yielded no significant effects of sentence type in any pre-critical region (F_1s < 1.970, $ps > .16$; F_2s < 2.717, $ps > .10$) in ambiguous sentences. In other words, we did not observe any consistent structurally driven differences in rereading times of pre-critical regions that would be predicted by selective reanalysis. In unambiguous controls, participants spent more time rereading the second noun phrase (NP) in GP sentences (Region 4; $M = 16.35$ ms) than in LC sentences (Region 3; $M = -119.57$ ms) [$F_1(1, 45) = 10.12$, $MSE = .228$, $p < .003$; $F_2(1, 36) = 7.75$, $MSE = .187$, $p < .008$]. There was also a significant interaction in this region [$F_1(1, 45) = 9.27$, $MSE = 4228$, $p < .004$; $F_2(1, 36) = 14.33$, $MSE = 4195$, $p < .001$] such that the ambiguity effect for GPs was smaller (effect = 125.96 ms) than the ambiguity effect for LCs (effect = 310.73 ms), which would be consistent with selective reanalysis predictions. However, given that we also just saw a sentence type difference in unambiguous sentences, it is not obvious that this latter interaction provides evidence for structurally driven rereading. No other significant differences in reading times were found in any pre-critical regions of the unambiguous controls (F_1s < 0.480, $ps > .50$; F_2s < 3.3, $ps > .08$).

Spillover region

The measures reported above all point to greater reading disruption caused by LC structures than the analogous GPs. Perhaps, however, due to the

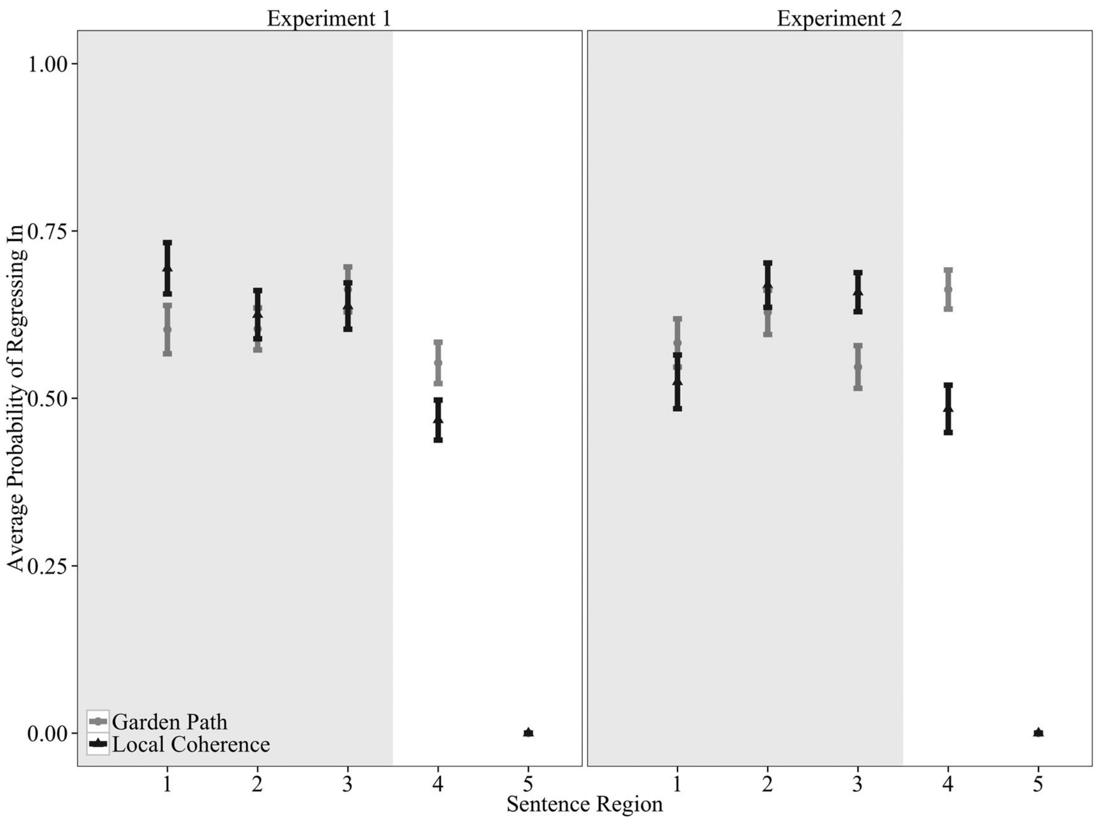

Figure 1. Probability of regressing into each sentence region (ambiguous trials only) for Experiments 1 (left panel) and 2 (right panel). Pre-critical regions are shaded. Note that probabilities are 0 for sentence-final regions. To view this figure in colour, please visit the online version of this Journal.

uncertainty caused by the fact that multiple structures are possible early in the GP RR structures, disruption simply lagged behind that in the LC RR structures. In other words, perhaps GP RRs caused equally as much trouble, but signs of it appeared later than the critical region reported above. To test this "delayed disruption hypothesis", we examined first-fixation duration, gaze duration, go-past time, and total time for the spillover region (the final region in the sentence). As can be seen in the bottom half of Table 3, results revealed a continued penalty for the LCs, with no effects on first-fixation duration (except for an item-level interaction), and a main effect of ambiguity for gaze duration, with shorter gaze durations in ambiguous conditions. This is probably due to the increased rereading that is done in the ambiguous conditions (inflated go-past times on the critical regions). There were no other effects or interactions in gaze duration. For both go-past time and total time, there were main effects of ambiguity and sentence type, with both effects patterning in the same way: Ambiguous conditions were longer than unambiguous, and LCs were longer than GPs. Analyses on raw (unresidualized) go-past times for this region replicated the analyses of residualized times: Ambiguous sentences led to longer go-past times ($M = 2836$ ms) than unambiguous sentences ($M = 2306$ ms). LC sentences led to longer go-past times ($M = 2727$ ms) than GP sentences ($M = 2415$ ms).

There were no significant interactions. For probability of first-pass regression out of this region, there were main effects of sentence type by items and marginal effects by subjects and ambiguity by items and marginal effects by subjects, and a sentence type by ambiguity interaction by items only. This interaction was characterized by inflated probability of first-pass regression out in the ambiguous GP condition. This weak interaction is consistent with the view that GP RRs caused lingering difficulty. But, taken together, the data from the spillover region do not support the idea that GP RR disruptions became apparent only later in processing.

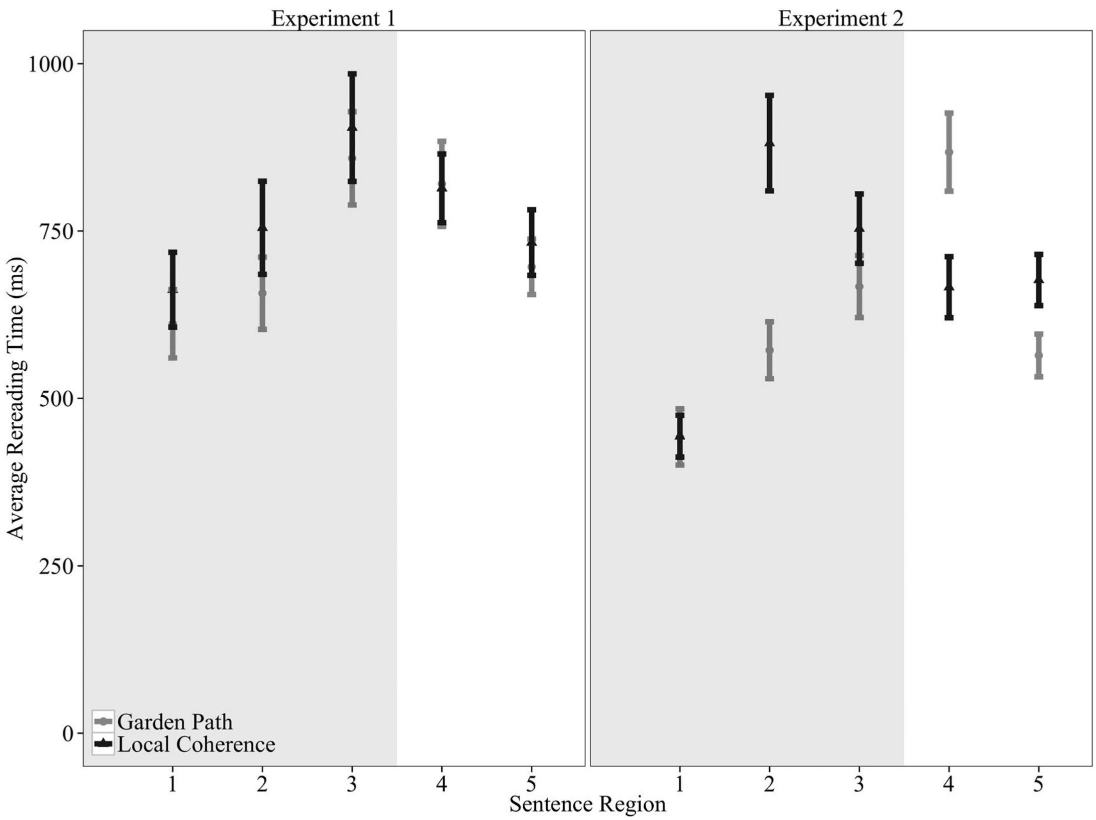

Figure 2. Rereading times in each sentence region (ambiguous trials only) for Experiments 1 (left panel) and 2 (right panel). Pre-critical regions are shaded. To view this figure in colour, please visit the online version of this Journal.

Summary

There was a consistent and considerable advantage for unambiguous sentences over ambiguous sentences. First-fixation duration showed evidence of larger early disruption in LC sentences compared to GP sentences, an interactive effect that was also partially reflected in gaze duration. Go-past times were also inflated for LC sentences. Somewhat inflated reading times on the spillover region in ambiguous GPs suggested lingering difficulty compared to LCs. With respect to measures related to rereading—first-pass regressions out of the critical region and into pre-critical regions, along with second-pass reading times on pre-critical regions—LC RRs triggered more regressions into Region 1 only, despite the fact that this region is not implicated in the ambiguity in LC structures. Based on the online data alone, then, we would be tempted to describe LC RRs as being generally more difficult than the lexically equivalent GP RR sentences used here. The offline data have yet to be considered, however, and may alter that interpretation.

Offline data

For the offline data, trials in which participants took longer than 10,000 ms to respond to the question were discarded (<1% of the data). Question response times were analysed using ANOVAs in the manner described for the online data. Mean question response times can be found in the left panel of Figure 3. Question response accuracy was analysed using logit mixed modelling (Jaeger, 2008).

Question response time

For question response time, there was a main effect of sentence type [$F_1(1, 44) = 26.98$, $MSE = 206{,}771$, $p < .0001$; $F_2(1, 35) = 26.35$, $MSE = 162{,}034$, $p < .0001$], indicating that questions about GP sentences were answered more quickly. Neither the main effect of ambiguity [$F_1(1, 44) = 1.5$, $MSE = 259{,}284$, $p = .23$; $F_2(1, 35) = 0.77$, $MSE = 146{,}013$, $p = .38$] nor the interaction was significant (all Fs < 1.5).

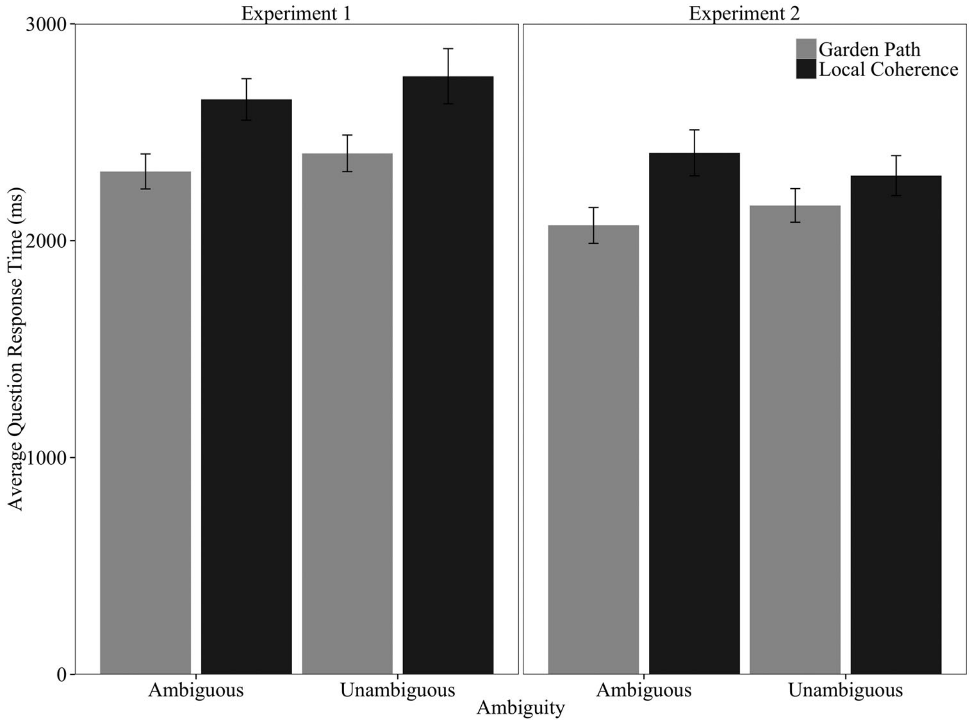

Figure 3. Question response time results for Experiments 1 (left panel) and 2 (right panel). To view this figure in colour, please visit the online version of this Journal.

Question response accuracy

A logit mixed model was used to analyse the binary question response accuracy data (Jaeger, 2008). The random effects structure was fitted using likelihood ratio rests and included random intercepts for participant and item and random by-participant and by-item slopes for sentence type. The model was fitted using a stepwise selection procedure, and only those effects that were significant or marginal were retained in the model. The final, best fitted model included sentence type and ambiguity, as well as the interaction of the two. The effect of ambiguity was highly significant (coefficient = 2.44, $SE = 0.18$, $z = 13.77$, $p < .0001$), indicating that questions about unambiguous sentences were responded to significantly more accurately than questions about ambiguous sentences. The effect of sentence type was also significant (coefficient = 1.79, $SE = 0.22$, $z = 8.2$, $p < .0001$), with higher accuracy for LC sentences than for GP sentences. A significant interaction of these two predictors (coefficient = −1.81, $SE = 0.23$, $z = -7.78$, $p < .0001$) indicates that the effect of sentence type was only significant for ambiguous sentences; accuracy was the same for questions about unambiguous GP and LC sentences. The left panel of Figure 4 illustrates this interaction. The effect of the ambiguity manipulation on accuracy is about 25% (11% for LC sentences and 39% for GP sentences), while the difference between the LC and GP conditions is about 29% for ambiguous sentences and a non-significant 1.4% for unambiguous sentences. (Note that the left panel of Figure 4 is presented using means for ease of exposition only, as logit models operate over individual responses rather than means.)

Relationship between regressions into pre-critical regions and accuracy

Schotter et al. (2014) demonstrate that regressions are positively related to comprehension accuracy: When readers are allowed to regress to previous text, comprehension is better than when readers are not allowed to regress, as in the trailing mask paradigm used by Schotter et al. However, comprehension on trials in that study in which readers regressed did not differ as a function of sentence ambiguity. In other words, regressions did not differentially improve comprehension of ambiguous sentences, which would be expected if those regressions were

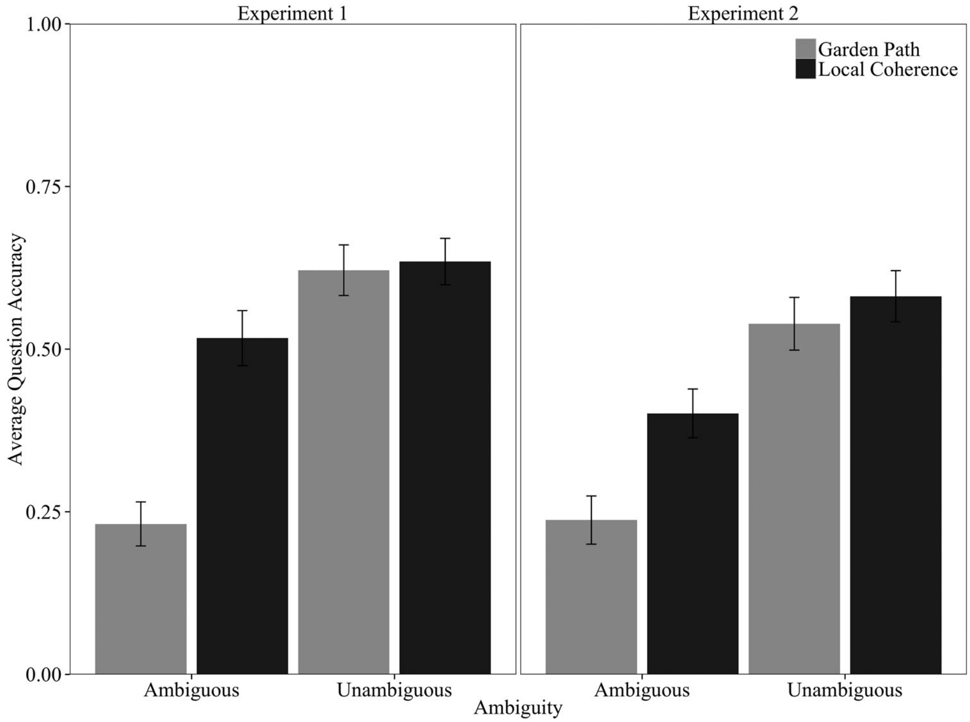

Figure 4. Question accuracy results for Experiments 1 (left panel) and 2 (right panel). To view this figure in colour, please visit the online version of this Journal.

launched in order to reanalyse previous structural commitments. We implemented the analysis employed by Schotter et al. to examine whether regressions into any particular pre-critical region were associated with better comprehension of ambiguous GPs and LCs. GLMER models of accuracy (binomial variable) included fixed effects of probability-of-regressing-in and sentence type, and random effects of subjects and items. A separate model was run for each pre-critical region. Probability-of-regressing-in at any region did not reach significance as a predictor of accuracy (all *p*s > .32). In Regions 1, 2, and 3, we found no interactions of probability-of-regressing-in with the effect of sentence type to predict accuracy (all *p*s > .13).

Discussion of Experiment 1

The combined results of Experiment 1 can be summarized succinctly: First-pass reading measures showed classic effects of temporary ambiguity, and comprehension accuracy was better for unambiguous sentences than for ambiguous sentences. On ambiguous trials, comprehension accuracy was better for LC RRs than for GP RRs, and the probability of regressing into Region 1 was greater in LC RRs than in GP RRs. However, the probability of regressing into any pre-critical region was not a predictor of offline comprehension accuracy. Furthermore, no second-pass time measures in any pre-critical region were related to comprehension accuracy (*p*s > .22). Finally, comprehension accuracy was not related to total time spent reading the critical region of ambiguous sentences (*p*s > .13). No prediction based on selective reanalysis was borne out: Regressions were not launched differentially into structurally important regions, nor was extra attention devoted to structurally important regions during rereading. Finally, no connection between rereading measures and comprehension accuracy was found.

The results of Experiment 1 appear to be more consistent with confirmatory, as opposed to revisionary, rereading. Specifically, if rereading is often more confirmatory than revisionary, then more time spent rereading LC RRs—in which early structural commitments need to be confirmed—should result in better comprehension. Confirmatory rereading in GP RRs, however, should result in worse comprehension, as earlier structure should be revised rather than confirmed. Inflated go-past times on the critical and spillover regions in the LC RRs compared to GP RRs were consistent with this confirmatory view of rereading; however, we found no relationship between where participants regressed to or how much time they spent rereading any particular region and their comprehension accuracy. Moreover, a purely confirmatory motivation for rereading would predict that readers should return to Region 1 equally often in both sentence types, if the main purpose of rereading is to check the contents of what had been read. Yet we found a higher probability of regression into Region 1 for LC RRs than for GP RRs. Therefore, confirmatory rereading cannot be concluded to be the only type of rereading that was done in these items.

Nevertheless, the combination of faster response times and lower accuracy for comprehension questions in the GP RR condition is suggestive of largely confirmatory rereading. Confirmation of the initial part of GP RRs returns the wrong parse (again), reinforcing it and supporting a globally incorrect parse. Confirmation of the initial part of LC RRs returns the correct parse, perhaps helping only nominally with integrating the later RR. If faster response times to comprehension probes signal relative confidence in or certainty of one's interpretation (Logačev & Vasishth, 2016), then confirmatory reading of GP RRs would be expected to lead to fast responses at below-chance accuracy levels. Confirmatory rereading of LC RRs, on the other hand, just brings the parser back to its ongoing struggle to integrate the ill-fitting incoming material, which it is only successful at doing about half the time.

Experiment 2

To ensure that the lack of structurally driven regressions and rereading and the low comprehension levels in Experiment 1 were not flukes, we replicated Experiment 1 using materials that were designed to eliminate one potential additional "local coherence" in the GP structures that were used in Experiment 1. Specifically, in (4a), the substring *the ball interfered with* could also be mistaken for a (second) reduced relative. Perhaps this added layer of temporary ambiguity was to blame for the low accuracy rates, although it would not explain the differences in go-past times on the disambiguating region. To eliminate this concern, Experiment 2 used materials exemplified in (5), which were slightly altered from Experiment 1 by including *had* prior to the main verb.

(5) a. The player | tossed | the ball | *had interfered with* | the other team.
b. The player | who was | tossed | the ball | *had interfered with* | the other team.
c. The other team | had interfered with | the player | *tossed* | the ball.
d. The other team | had interfered with | the player | who was | *tossed* | the ball.
e. Did the player toss the ball?

Method

Participants

Forty-eight people from the University of Illinois at Urbana-Champaign community participated. The large majority of participants were recruited from the Educational Psychology subject pool. All participants were compensated for their time with either course credit or $7.

Apparatus, materials, and procedure

The apparatus and materials were identical to those in Experiment 1, with the exception that the critical items followed the construction similar to that of (5). In addition, the 40 experimental items were interleaved among 160 other sentences with a variety of structures, all of which were followed by comprehension probes of some sort. Eighty percent of these other sentences were items from unrelated experiments, including other types of garden paths (direct object/sentence; conjunction) and long-distance agreement. Twenty percent were true fillers, consisting of unambiguous complex sentences. Like Experiment 1, the correct answer to 50% of the items across all item types was Yes; that to the other 50% was No. Mean accuracy on true filler sentences in Experiment 2 was 89%; mean accuracy for all items excluding the items for this experiment was 72% (range: 44%–90%). The procedure was identical to that in Experiment 1. Each trial lasted no more than 60 minutes.

Results

Online data

Descriptive statistics for raw reading times and probability of first-pass regression out of the critical region are reported in Tables 1 and 2. Residualized reading times were used in the analyses, as described for Experiment 1. The critical region for each sentence type was the same as in Experiment 1: In the GP conditions, this was the disambiguating verb (*had interfered with*). In the LC conditions, this region was the relative clause verb (*tossed*), which was locally coherent with the direct object of the main verb (*player*) in the ambiguous condition.

Identical to Experiment 1, the analysis of the online data in the critical and spillover regions included five different dependent measures of reading time were analysed using repeated measures ANOVAs: first-fixation duration, gaze duration, go-past time, total time, and first-pass regressions out. We also conducted analyses of probability of regressions into and second-pass time for the pre-critical regions. For each ANOVA, list condition was included as a between-participants and between-items factor (Pollatsek & Well, 1995) to remove variance associated with the individual lists. The effects from all ANOVAs can be found in the right panels of Table 3. Before analysis, fixations shorter than 80 ms and longer than 1200 ms were excluded from the data set (6.91% of fixations).

Critical region

First-fixation duration. For first-fixation duration, there was a main effect of sentence type. The main effect of ambiguity was not significant, but ambiguity and sentence type interacted. Follow-up *t* tests indicated that the effect of sentence type was only significant for the ambiguous sentences [$t_1(47) = -4.418$, $p < .0001$; $t_2(39) = -4.015$, $p = .0003$], and not for the unambiguous sentences (both $ts < 1.48$). As in Experiment 1, ambiguous LC sentences caused a significant early disruption to reading.

Gaze duration. The main effect of sentence type was significant, indicating longer reading times on GP items. The main effect of ambiguity was also significant, indicating longer reading times for ambiguous sentences. The interaction was not significant.

Go-past time. For go-past time, there was a main effect of ambiguity, with inflated go-past times in ambiguous conditions. There was no main effect of sentence type and no interaction. Analyses of raw (uresidualized) go-past times on the critical region replicated the analyses of residualized go-past times: Ambiguous items led to longer go-past times ($M =$ 781 ms) than unambiguous sentences ($M = 603$ ms).

Total time. For total time, the main effect of sentence type was not significant. The main effect of ambiguity was significant, indicating longer fixations for ambiguous sentences. The interaction was significant. Follow-up *t* tests indicated that the effect of sentence type was only significant for the ambiguous sentences [$t_1(47) = 2.826$, $p < .0069$; $t_2(39) = 2.442$, $p = .019$], and not for the unambiguous sentences (both $ts < .60$). The interaction was driven by the inflated total reading times in the ambiguous GP condition.

Probability of first-pass regressions out of the critical region. Analyses of the probability that a regression would be launched from this region to a previous part of the sentence revealed no main effect of sentence type. There also was no main effect of ambiguity. There was an interaction of sentence type and ambiguity by items only such that first-pass regressions out were more likely for ambiguous sentences only in the GP condition.

Pre-critical regions

Probability of regressing into pre-critical regions. Consistent with the analysis in Experiment 1, the probability of regressions into pre-critical regions was analysed for each pre-critical region of ambiguous items only. Recall that the selective reanalysis hypothesis (Frazier & Rayner, 1982) predicts that readers should direct regressions differentially in GPs and LCs. Specifically, in ambiguous GPs, Region 1 (*The player*) and Region 2 (*tossed*) should be targeted for regressions. In ambiguous LCs, Region 3 (*the player*) should be targeted as the beginning of the reduced relative, if regressions are initiated at all (cf. Mitchell et al., 2008). The right panel of Figure 1 illustrates the mean probabilities for regressions into these (shaded) pre-critical regions.

There was a marginal effect of sentence type in Region 1 [$F_1(1, 44) = 3.53$, $MSE = .081$, $p = .067$; $F_2(1, 36) = 2.89$, $MSE = .056$, $p = .098$], in the same direction as that in Experiment 1 (LC > GP), and no effects in Region 2 ($ps > .18$). However, we did find a main effect of sentence type in Region 3 [$F_1(1, 44) = 20.88$, $MSE = .300$, $p < .0001$; $F_2(1, 36) = 7.50$, $MSE = .224$, $p = .009$], such that there were more regressions into Region 3 in LC RRs than in GP RRs. This difference in Region 3 is expected under selective reanalysis.

In unambiguous controls, we observed higher probabilities of regression into ambiguous structures (compared to unambiguous items) for both the relativizer (*who was*) [$F_1(1, 44) = 4.81$, $MSE = .145$, $p = .037$; $F_2(1, 36) = 6.88$, $MSE = .135$, $p = .013$] and the verb [$F_1(1, 44) = 5.98$, $MSE = .229$, $p = .019$; $F_2(1, 36) = 6.97$, $MSE = .200$, $p = .012$], with LCs again triggering more rereading than GPs. We also observed significant interactions between sentence type and ambiguity in Region 1 [$F_1(1, 44) = 7.48$, $MSE = .128$, $p < .009$; $F_2(1, 36) = 9.44$, $MSE = .115$, $p < .004$] and Region 3/4 (second NP) [$F_1(1, 44) = 12.96$, $MSE = .283$, $p < .001$; $F_2(1, 36) = 8.79$, $MSE = .207$, $p < .006$]. The first interaction was characterized by a smaller ambiguity effect for GPs (effect = .220) than for LCs (.374) at the initial noun phrase, while the second interaction was due to a larger ambiguity effect for GPs (effect = .114) than for LCs (.011) at the second noun phrase. This pattern is opposite to what might be expected under selective reanalysis.

We also examined rereading times in the pre-critical regions to look for evidence that increased regressions into pre-critical regions were associated with inflated second-pass times in those regions.

Rereading times on pre-critical regions. Rereading times reflect time spent rereading regions after already having left them to the right. Like regressions in, selective reanalysis predicts that second-pass times should be related to reanalysis and recovery processes. Unnormalized rereading times by region are illustrated in the shaded portion of the right panel of Figure 2. In Region 1, there was no main effect of sentence type ($ps > .11$).

In Region 2, there was a main effect of sentence type ($F_1(1, 44) = 16.55$, $MSE = 730{,}999$, $p < .0001$; $F_2(1, 36) = 18.72$, $MSE = 935{,}959$, $p < .0001$]; LCs were read more slowly in Region 2 than GPs.

The same pattern held in Region 3 [$F_1(1, 44) = 9.77$, $MSE = 232{,}193$, $p = .003$; $F_2(1, 36) = 10.23$, $MSE = 293{,}722$, $p = .003$], again with LCs read more slowly than GPs. Like regressions into this region, the inflated rereading times for LC RRs in Region 3 are consistent with selective reanalysis, but the inflated rereading times in Region 2 are not.

With respect to reading times on pre-critical regions in unambiguous controls, there were interactions on Region 1 [$F_1(1, 40) = 3.27$, $MSE = 1199$, $p < .08$; $F_2(1, 36) = 10.21$, $MSE = 0.2231$, $p < .003$], Region 2 [$F_1(1, 43) = 7.73$, $MSE = 3357$, $p = .008$; $F_2(1, 36) = 6.89$, $MSE = 2762$, $p < .013$], and Region 3/4 [the second NP; $F_1(1, 42) = 18.39$, $MSE = 4636$, $p < .001$; $F_2(1, 36) = 11.16$, $MSE = 3600$, $p < .002$]. In the first noun region, the interaction was such that ambiguity effects were larger for GPs (effect = 96.55 ms) than LCs (effect = 11.77 ms). The opposite was true of the interactions in the remaining two regions, such that LCs elicited larger ambiguity effects than GPs (Region 2: GP effect = 109.36 ms, LC effect = 270.34 ms; Region 3/4: GP effect = 98.60 ms, LC effect = 296.53 ms). This pattern was more consistent than that in Experiment 1 and suggests that participants tended to reread the first NP in ambiguous GP sentences, as might be expected under a broad application of selective reanalysis. On the other hand, all subsequent regions were reread more in ambiguous LC sentences, which is not consistent with selective reanalysis.

In Experiment 2, some of the predictions of selective reanalysis fared better than they had in Experiment 1. Region 3 was regressed into more in LC RRs and was also reread longer in LC RRs. However, Region 2 also received more rereading in LC RRs, which is opposite of what is predicted by selective reanalysis. In first-pass reading measures, the inflated first-fixation durations on LC RRs observed in Experiment 1 were replicated. On the other hand, GP RRs showed inflated gaze durations and total times in Experiment 2, opposite Experiment 1, yet there was no sentence type effect in go-past times, whereas in Experiment 1, LC RRs triggered longer go-past times.

Spillover region

Once again, to examine the effects of potential uncertainty caused by the fact that multiple structures are possible early in the ambiguous GP sentences, we examined first-fixation duration, gaze duration, go-past time, total time, and probability of first-pass regressions out of the region following the critical region (from the word following the critical verb to the end of the sentence). The bottom panels of Tables 1 and 2 show these descriptive statistics. Results revealed significant sentence type by ambiguity interactions on first-fixation duration and gaze duration, but in the opposite direction of the "delayed disruption" hypothesis—that is, inflated durations on these measures in LC RRs compared to GP RRs. No main effects reached significance. Thus, as in Experiment 1, early measures showed a continued penalty for LC structures (see Tables 1 and 2). There were main effects of ambiguity and sentence type on both total time and go-past time, as in Experiment 1, with both effects patterning in the same way: Ambiguous conditions were longer than unambiguous, and LCs were longer than GPs. Analyses of raw (unresidualized) go-past times replicated the analyses of residualized go-past times: Ambiguous sentences led to longer go-past times ($M = 2360$ ms) than unambiguous sentences ($M = 1710$ ms). LC sentences led to longer go-past times ($M = 2195$ ms) than GP sentences ($M = 1876$ ms). There were no significant interactions on these later measures. For probability of first-pass regressions out of this region, there was a sentence type by ambiguity interaction. This interaction was characterized by an inflated probability of first-pass regressions out in the ambiguous GP condition, as in Experiment 1. Taken together, the data from the spillover region do not support the idea that GP-related disruptions became apparent later in processing than LC disruptions.

Offline data

For the offline data, trials where participants took longer than 10,000 ms to respond to the question were discarded (<0.25% of the data). Question response times were analysed using ANOVAs in the manner described for the online data. Average question response time can be found in the right panel of Figure 3. Question response accuracy was analysed using logit mixed modelling (Jaeger, 2008), and these are illustrated in the right panel of Figure 4.

Question response time

For question response time, there was a main effect of sentence type [$F_1(1, 44) = 18.96$, $MSE = 2,625,757$, $p < .0001$; $F_2(1, 35) = 12.52$, $MSE = 1,985,090$, $p = .001$], indicating that questions about GP sentences were answered more quickly. There was no main effect of ambiguity (both Fs $< .029$). The interaction was significant by participants but only marginally by items [$F_1(1, 44) = 4.173$, $MSE = 514,205$, $p = .047$; $F_2(1, 35) = 2.953$, $MSE = 440,531$, $p = .095$]. Follow-up t tests indicated that the effect of sentence type was significant for the ambiguous sentences [$t_1(47) = -4.03$, $p = .0002$; $t_2(39) = -3.34$, $p = .002$], and marginally significant for unambiguous sentences by participants but not by items [$t_1(47) = -1.87$, $p = .068$; $t_2(38) = -1.44$, $p = .158$].

Question response accuracy

A logit mixed model was used to analyse the binary question response accuracy data (Jaeger, 2008). The random effects structure was fitted using likelihood ratio rests and included random intercepts for participant and item and random by-participant and by-item slopes for sentence type. The model was fitted using a stepwise selection procedure, and only those effects that were significant or marginally so were retained in the model. The final, best fitted model included sentence type and ambiguity, as well as the interaction of the two. The effect of ambiguity was highly significant (coefficient = 0.287, $SE = 0.02$, $t = 12.946$, $p < .0001$), indicating that questions about unambiguous sentences were responded to significantly more accurately than questions about ambiguous sentences. The effect of sentence type was also significant (coefficient = 0.12, $SE = 0.02$, $t = 5.48$, $p < .0001$), with higher accuracy for LC sentences than for GP sentences. The interaction of these two predictors (coefficient = −0.23, $SE = 0.04$, $t = -5.27$, $p < .0001$) indicates that the effect of sentence type was only significant for ambiguous sentences; accuracy was the same for questions about unambiguous GP and LC sentences. The right panel of Figure 4 illustrates this interaction. The effect of the ambiguity manipulation on accuracy is about 24% (18% for LC sentences and 31% for GP sentences), while the difference between the LC and GP conditions is about 17% for ambiguous sentences and a non-significant 3% for unambiguous sentences. (Note again that Figure 4 is presented using means for ease of exposition only.)

Relationship between regressions into pre-critical regions and accuracy

As in Experiment 1, we extended the analysis employed by Schotter et al. (2014) to examine whether regressions into any particular pre-critical region were associated with better comprehension of ambiguous GPs and LCs only. GLMER models of accuracy (binomial variable) included fixed effects of probability-of-regressing-in and sentence type, and random effects of subjects and items. A separate model was run for each pre-critical region. Probability-of-regressing-in to any region did not reach significance as a predictor of accuracy (all ps $> .18$), nor did regressions in interact with sentence type to predict accuracy (all ps $> .12$). In other words, the likelihood of readers regressing into any pre-critical region—irrespective of whether that region was implicated in the temporary ambiguity—bore no relationship to comprehension accuracy. There were also no significant effects of sentence type or rereading time on ambiguous sentences (ps $> .15$), nor interactions of these two factors (ps $> .07$). Finally, there was no relationship between total time spent on the critical region of ambiguous sentences and comprehension accuracy (ps $> .51$).

Summary

There was a consistent and considerable advantage for unambiguous (unreduced) sentences over ambiguous (RR) sentences. The effect of sentence type varied across online and offline measures. First-fixation duration and response times to questions showed evidence of larger processing disruptions in LC sentences than in GP sentences. This pattern reversed for gaze duration and total time and for comprehension accuracy, such that GP sentences were processed more slowly and led to poorer comprehension accuracy with shorter response latencies than LC structures. Where participants regressed to and how long they spent rereading regions of the sentences bore no discernible relation overall to their comprehension accuracy, which was, once again, very low, although comprehension was again better overall for LC structures. What might be targeted regressions into and more concerted rereading of Region 3 in LC RRs, as predicted by selective reanalysis, seem to be contradicted by similar results for Region 2, not predicted by selective reanalysis. Moreover, none of the rereading behaviour predicted by selective reanalysis for GP RRs was observed. As in Experiment 1, no connection was found between rereading and comprehension accuracy. Furthermore, the lack of a sentence type difference in regressions into Region 1 (a difference that was observed in Experiment 1) suggests no structural difference in the likelihood of rereading the entire sentence, as predicted under a more confirmatory view of rereading. In summary, although some of the eye movement measures were consistent with selective reanalysis, the lack of relationship between these and offline comprehension measures raises the question of whether reanalysis was actually performed during rereading.

General discussion

This study is the first to compare garden-path main verb/reduced relative structures (GP RRs) and local coherence reduced relative structures (LC RRs) head-to-head, and to demonstrate their relative difficulties using both online eye movement measures and offline comprehension measures. Our results suggest that both structures pose extreme difficulties for the language processor, but they differ in theoretically interesting ways. The offline results are quite striking: Comprehension accuracy rates were only about 60% even for unambiguous versions, and the GP RRs were comprehended correctly only about 25% of the time. Comprehension accuracy for the LC RRs was basically at chance across the two experiments. It might be argued that participants were not reading carefully; however, high accuracy rates on non-experimental items in both experiments suggest otherwise. Alternatively, perhaps a "yes bias" depressed levels of accurate "No" responses. If this explanation for the absolute accuracy levels is accurate, it leaves the significant relative effects of sentence type and ambiguity unaffected: LC RRs were comprehended more accurately than GP RRs (consistent with Gibson, 1998, 2000), and unambiguous sentences were comprehended more accurately than ambiguous sentences, irrespective of sentence type (contra Gibson, 1998, 2000). We note also that the eye-movement patterns elicited by these sentences were typical of those reported in previous eye-tracking experiments with garden-path sentences. These eye movements were interpreted in that previous work as indices of reanalysis (e.g. Meseguer et al., 2002) In the experiments reported here, though, we found no obvious relationship between where regressive eye movements were directed nor the amount of time spent rereading and comprehension accuracy.

Walking through the eye movement results, we focus on the rereading-related measures. According to the selective reanalysis hypothesis (Frazier & Rayner, 1982), participants should have targeted different regions of GP RRs and LC RRs for rereading. Importantly, if rereading is performed in order to reanalyse earlier misparsed structure, and thereby recover from misanalysis and/or misinterpretation, regressive eye movements in GP RRs should have mainly targeted the first two regions of the sentence. In contrast, regressive eye movements in LC RRs, if launched at all, should have mainly targeted the region immediately preceding the critical region (Region 3). Moreover, if rereading can confidently be assumed to index *successful* recovery from temporary ambiguities (GP or LC), then there should be some observable connection between what is reread, how long it is reread, and ultimate comprehension accuracy.

Contrary to these predictions, we observed no clear relationship between any measures of rereading and comprehension accuracy. Especially in Experiment 2, there were scattered signs that readers were directing regressions back to some structurally relevant regions (but not others); however, their doing so did not improve the likelihood that they would answer comprehension questions accurately. The only consistent finding regarding accuracy in ambiguous conditions was that participants comprehended the LC RRs more accurately than the GP RRs. Response times were also consistently longer for LC RRs than for GP RRs.

It might be proposed that LC RRs were not in fact easier to comprehend than the GP RRs. Perhaps the inflated response latencies to the comprehension probes for the LC RRs reflect a total parsing failure or "parsing panic" (a term suggested by an anonymous reviewer). It is true that, anecdotally, LC RRs "feel" stranger than GP RRs, at least to psycholinguists who have been reading and creating GP RRs at least 20 years longer than LC RRs. Fine, Jaeger, Farmer, and Qian (2013) demonstrate that even short exposure can make difficult sentence structures seem more acceptable and easier to process, however, so the intuitions of researchers who are relatively more habituated to GP RRs than LC RRs might not be reliable. Additionally, if we accept parsing panic as an alternative explanation for the LC results, we are left to wonder why the regressions in the GP RRs were equally widely dispersed across regions: Would "panic" not be expected to trigger frantic searching for information compared to a more "selective" reanalysis process (i.e., in the presumably non-panic-inducing GP RRs)? Individual differences in responses to both GP RRs and LC RRs are quite likely, however, and it could well be that some readers "panic", while others reread to confirm, and still others reread to revise. Furthermore, the same reader might pursue any one of these strategies, depending on the task at hand. Individual differences in rereading and cognitive resources have been demonstrated by von der Malsburg and Vasishth (2011, 2013). Our present experimental design does not allow us to investigate these individual differences, but continued research along these lines is underway. A fruitful continuation of this line of investigation should include different types of comprehension probes, including forced-choice questions ("Who tossed the ball?"), paraphrases (cf. Patson et al., 2009), and recall tasks.

On grounds of parsimony then, we interpret the results in the same way for both constructions. The lack of connection between rereading measures and accuracy and the sentence type difference in accuracy suggest that rereading was generally more confirmatory than revisionary, whether confirmatory rereading was appropriate or not. When readers encounter unexpected material or material that does not fit into the unfolding parse, they seem uncertain about what they have read (Levy et al., 2008) and launch regressions to sample previous material. Some of these regressions appear to be structurally targeted; some, however, appear to be relatively arbitrary (cf. Mitchell et al., 2008). In either case, the subsequent rereading seems to be done largely to confirm what was previously read. In the case of GP RRs, what was previously read was incorrectly parsed, and it appears to be incorrectly parsed again upon a confirmatory rereading. Lewis (1998) noted that rereading of all or a large portion of a sentence might be pursued as a reanalysis strategy in order to save on memory. He observed that following this strategy would require the reader only to remember not to go down the same incorrect path again. The present data suggest, however, that readers did in fact go down the same path again when they reread GP RRs.

In the case of LC RRs, what had been previously read was correctly parsed, so "going down the same path" was warranted. As such, attention could be focused on figuring out how the locally coherent string could be integrated. It is important to point out that the amount of confirmatory rereading in LC RRs is unsurprisingly *not* related to ultimate comprehension accuracy, as rereading early regions does not entail any reanalysis. This integration process is more likely to be successful than the restructuring process required for accurate comprehension of the GP RR, although it is not overly successful. If this characterization is accurate, then the "selective reanalysis hypothesis" might be better termed the "selective *regression* hypothesis", as regression, even to structurally targeted regions, does not seem to imply reanalysis, recovery, or accurate comprehension.

An alternative explanation along the lines of "noisy channel" models discussed above (e.g. Gibson et al., 2013; Levy, 2008; Levy et al., 2009) might also be considered. Levy et al. (2009) argued that people's memories for previous text is fallible, and when they encounter difficult text, they might mentally "edit" previous text or "repair" a perceived typographical error. In the case of the GP RRs and LC RRs, adding an *and* between clauses could lead to the observed pattern of comprehension results, as in (6).

(6) a. The player tossed the ball *and* interfered with the other team. (GP)
b. The other team interfered with the player *and* tossed the ball. (LC)

Although some of the rereading measures on LC RRs suggest that readers focused on the regions where an "edit" would be made, this is not true of the measures on GP RRs. No explicit predictions about the link between editing processes and regressions or rereading have been made within these models, as far as we know. In any case, these edits would remain examples of readers arriving at final interpretations that are unfaithful to the input, despite having had unlimited time to read and reread the input. Rereading then could be described as "revisionary" after all, rather than "confirmatory", but not in the way as that predicted by the selective reanalysis hypothesis.

We therefore feel that the pattern of results here is most consistent with previous work on good-enough processing, showing that readers do not always arrive at veridical representations of the actual content of the input (Christianson et al., 2001, 2006; Ferreira, 2003; Ferreira et al., 2002; Ferreira et al., 2001; Ferreira & Patson, 2007; Slattery et al., 2013). The body of research within the good-enough framework has suggested that although reanalysis of garden-path structures might not run to completion —or be undertaken at all—readers do not seem to perceive comprehension failure or general confusion (at least not college-aged readers; cf. Christianson et al., 2006). In fact, readers correctly answer questions about the main clauses of garden-path sentences like *While Anna dressed the baby spit up on the bed* (e.g. *Did the baby spit up?*) while incorrectly answering questions about the subordinate clause (*Did Anna dress the baby?)*, both with high degrees of confidence. In instances where a plausible semantic interpretation can be formed without figuring out the correct global syntax, readers show a tendency to do so. Previous studies have provided evidence that reanalysis is a "last resort" (Sturt et al., 2001) and that the parser is loath to relinquish a plausible misparse (Christianson et al., 2001, 2006; Ferreira & Henderson, 1991). The current results suggest that readers are slower to register the error signal in garden paths (though this slowness does not extend into the spillover region), generally spend less time trying to integrate incoming information with the initial portion of the parse, fail to revise that initial portion (where necessary), and strongly maintain the initial interpretation. Alternatively, along the lines of Slattery et al. (2013), the initial structure is revised, but the initial interpretation is nevertheless maintained (perhaps concurrently with correct interpretation); however, if the two interpretations here were being maintained and thus were in competition with one another, the garden-path accuracy rates should have been nearer chance. In the LC RRs, more effort is generally devoted to trying to fit incoming discordant material into previous unambiguous structure. However, across the two experiments, regressions in LC RRs appear to be scattered across Regions 1 (Experiment 1) and Regions 2 and 3 (Experiment 2). This variability is inconsistent with selective reanalysis, but is quite consistent with individual differences in scanpaths observed by von der Malsburg and Vasishth (2011, 2013). Specifically, in their 2013 paper, they found that readers with higher working memory capacities tended to make structural commitments during the first pass through garden-path sentences and to be more likely to go back and reread. Therefore, any future investigations into the connection between regressions, rereading, and offline comprehension should include individual difference measures. In any case, the combined extra rereading effort in LC RRs did appear to result in a better outcome, in as much as near-chance performance can be characterized as "better".

The extremely low offline accuracy rates that we observed are of theoretical significance. Since the groundbreaking work of Frazier and Rayner (1982, 1987), it has been widely accepted that rereading-related measures index "recovery" processes. And indeed, in the present experiment, we observed inflated rereading measures (regressions, second-pass times, and go-past times) for "ambiguous" versus "unambiguous" GP and LC structures, consistent with over three decades of eye-tracking research on syntactic parsing. However, these regressions and rereading efforts were not terribly useful in helping participants arrive at the correct interpretation of these sentences. To the extent that a faster response can be interpreted as a sign of certainty (e.g. Logačev & Vasishth, 2016), overall less time devoted to rereading GP RRs appears to have led to a greater degree of certainty in inaccurate responses to comprehension probes. In contrast, relatively more time spent rereading LC RRs led to less certainty (or guessing) and relatively more accurate responses. Both of these patterns can plausibly be ascribed to confirmatory rereading of earlier material. In GP RRs, confirmatory rereading confirms the initial, ultimately incorrect parse, which is subsequently reinforced by the comprehension probe (cf. Christianson & Luke, 2011). In LC RRs, confirmatory rereading is not overly helpful in integrating the unexpected later RR. What, then, are we to take away from online measures that appear to show people dutifully rereading text and offline comprehension showing that they nevertheless misinterpreted the sentences 50–75% of the time (and ~35% of the time in *unambiguous* conditions)?

As proposed here, rereading may signal confirmation of previous interpretations and a cobbling together of several smaller interpretations based on substrings that are not fully integrated into a single global structure. As long as these sub-interpretations are basically compatible and coherent, they can be maintained simultaneously (Christianson et al., 2001, 2006). This interpretation of the results is broadly consistent with Frazier and Fodor's (1978) classic sausage machine model of syntactic parsing. In that two-stage model, the preliminary phrase packager was hypothesized to assign syntactic structure to short substrings, followed by the sentence structure supervisor, which took the parsed substrings and attached them together by adding non-terminal syntactic nodes to a global syntactic structure. It is possible that effects of the sort observed here derive from instances where the second stage is not completed, and the interpretations derived from the individual "sausage links" outputted by the preliminary phrase packager are cobbled together in a good-enough fashion—that is, in a final representation that is not faithful to the actual input (Ferreira et al., 2001, 2002; Ferreira & Patson, 2007; see also Christiansen & Chater, in press; Ferreira & Christianson, in press). In future work, interpretations could be probed in a way to determine whether representations from the first and second clauses of the sentences coexist simultaneously. Clearly, further research is needed to develop an explicit linking hypothesis between rereading measures, reanalysis, and interpretive processes, given that regressions and rereading in the two different structures compared here cannot index the same type of computations.

Acknowledgements

Portions of this work were presented at AMLaP 2011, Paris. We thank the audience of that presentation for helpful questions and discussion. We also are grateful for the excellent work of the members of the EdPsych Psycholinguistics Lab, Beckman Institute, especially Cassie Palmer.

Funding information

This work was supported by National Science Foundation (NSF) CAREER Award [grant number BCS-0847533] to K. Christianson and National Institutes of Health (NIH) Training [grant number T32-HD055272] (PI: Sarah Brown-Schmidt) to E. Hussey.

ORCID

Erika K. Hussey http://orcid.org/0000-0002-9782-5943

References

Abney, S. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research*, *18*, 129–144.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Amlund, J. T., Kardash, C. A. M., & Kulhavy, R. W. (1986). Repetitive reading and recall of expository text. *Reading Research Quarterly*, *21*, 49–58.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory*, *9*, 365–381.

Christiansen, M. H., & Chater, N. (in press). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-Enough, underspecified, or shallow language processing. *The Quarterly Journal of Experimental Psychology*, *69*, 817–828.

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. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*, *15*, 136–166.

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

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "Good-Enough" interpretations of garden-path sentences. *Discourse Processes*, *42*, 205–238.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–120.

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. (in press). Is now-or-never language processing Good Enough? Commentary on Christiansen & Chater. *Behavioral and Brain Sciences*.

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., Jr. (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., & Patson, N. D. (2007). The "Good Enough" approach to language comprehension. *Language and Linguistics Compass*, *1*, 71–83.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaption during syntactic comprehension. *PLOS One*, *8*(10), doi: 10.1371/journal.pone.0077661

Fodor, J. D., & Frazier, L. (1980). Is the human sentence parsing mechanism an ATN? *Cognition*, *47*, 247–275.

Frazier, L. (1998). Getting there (slowly). *Journal of Psycholinguistic Research*, *27*, 123–146.

Frazier, L., & Clifton, C. Jr. (1996). *Construal*. Boston, MA: MIT Press.

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.

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, *26*, 505–526.

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

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

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*, 8051–8056.

Gorrell, P. (1998). Syntactic analysis and reanalysis in sentence processing. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 201–246). Dordrecht: Kluwer Academic Publishers.

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

Hussey, E. K., Ward, N., Christianson, K., & Kramer, A. F. (2015, November 3). Language and memory improvements following tDCS of executive-control brain regions. *PLoS ONE*. doi:10.1371/journal.pone.0141417

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*, 15–47.

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.

Lewis, R. L. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In J. Fodor, & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 247–284). Boston, MA: Kluwer Academic Publisher.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Dordrecht: Elsevier Science.

Logačev, P., & Vasishth, S. (2016). Understanding underspecification: A comparison of two computational implementations *The Quarterly Journal of Experimental Psychology*, *69*, 996–1012.

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

Meseguer, E., Carreira, M., & Clifton, C. Jr. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*, 551–561.

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye- movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language*, *59*, 266–293.

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 280–285.

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, and Cognition*, *21*, 785–794.

Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C., Jr. (2012). *Psychology of reading* (2nd ed.). New York: Taylor & Francis.

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension Is supported by regressions during reading. *Psychological Science*, *25*, 1218–1226.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*, 104–120.

Spivey-Knowlton, M. J., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227–267.

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.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. Jr. (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.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 431–450.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehensions: The integration of habits and rules*. Cambridge, MA: MIT Press.

Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and Good-Enough parsing. *Trends in Cognitive Sciences*, *18*, 605–611.

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.

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 and Language*, *55*, 335–362.

von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, *65*, 109–127.

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*, 1545–1578.

Zhou, P., & Christianson, K. (2016). I "hear" what you're "saying": Auditory perceptual simulation, reading speed, and reading comprehension. *The Quarterly Journal of Experimental Psychology*, *69*, 972–995.

Appendix

Items: (Backslashes mark interest regions for analyses; they were not seen by participants. Odd-numbered items are garden-path structures; even-numbered items are the corresponding local coherence structures. The regions in parentheses "disambiguated" the items in the "unambiguous" conditions.)

(1) The babysitter /(who was) /purchased /a gift card /thanked /the parents.
(2) The parents /thanked /the babysitter /(who was) /purchased /a gift card.
(3) The boy /(who was) /sent /an invitation /hugged/ the tutor.
(4) The tutor /hugged /the boy /(who was) /sent /an invitation.
(5) The boyfriend /(who was) /left /a voicemail /distrusted /the girlfriend.
(6) The girlfriend /distrusted /the girlfriend /(who was) /left /a voicemail.
(7) The Brit /(who was) /served /the first ace /defeated /the Italian.
(8) The Italian /defeated /the Brit /(who was) /served /the first ace.
(9) The chef /(who was) /presented/ the new knife /greeted /the audience.
(10) The audience /greeted /the chef /(who was) /presented /the new knife.
(11) The child /(who was) /bought /an ice cream cone /smiled at /the cashier.
(12) The cashier /smiled at /the child /(who was) /bought /an ice cream cone.
(13) The child/ (who was) /read/ a story /hugged /the nanny.
(14) The nanny /hugged /the child /(who was) /read /a story.
(15) The child/ (who was) /saved /a front row seat /surprised /the clown.
(16) The clown /surprised /the child /(who was) /saved /a front row seat.
(17) The colonel/ (who was) /granted /leave /phoned /a travel agent.
(18) A travel agent /phoned /the colonel /(who was) /granted leave.
(19) The dispatcher /(who was) /handed /a pink slip /worried about /the driver.
(20) The driver /worried about /the dispatcher /(who was) /handed /a pink slip.
(21) The doctor /(who was) /delivered /the donor organ /assisted /several surgeons.
(22) Several surgeons /assisted /the doctor /(who was) /delivered /the donor organ.
(23) The general /(who was) /awarded /the medal /annoyed /the soldier.
(24) The soldier /annoyed /the general /(who was) /awarded /the medal.
(25) The girl /(who was) /passed /a note /tattled on /the classmate.
(26) The classmate / tattled on /the girl /(who was) /passed /a note.
(27) The girl /(who was) /played /a song /photographed /the tourist.
(28) The tourist /photographed /the girl /(who was) /played /a song.
(29) The girl /(who was) /won /a stuffed bear /ignored /the carnival worker.
(30) The carnival worker /ignored /the girl /(who was) /won /a stuffed bear.
(31) The golfer /(who was) /cooked /a special meal /toasted /the host.
(32) The host /toasted /the golfer /(who was) /cooked /a special meal.
(33) The hobo /(who was) /cut /a slice of cheese /despised /the beggar.
(34) The beggar / despised /the hobo /(who was) /cut /a slice of cheese.
(35) The inspector /(who was) /paid /a fee /argued with /the fireman.
(36) The fireman /argued with /the inspector /(who was) /paid /a fee.
(37) The instructor /(who was) /submitted /the plagiarized papers /visited /the dean.
(38) The dean/ visited /the instructor /(who was) /submitted /the plagiarized papers.
(39) The manager /(who was) /prepared /the report /praised /the employees.
(40) The employee /praised /the manager /(who was) /prepared /the report.
(41) The millionaire /(who was) /sculpted /a statue /bragged to /the banker.
(42) The banker /bragged to /the millionaire /(who was) /sculpted /a statue.

(43) The neighbour /(who was) /crocheted /a blanket /envied /the mailman.
(44) The mailman /envied /the neighbour /(who was) /crocheted /a blanket.
(45) The old man /(who was) /awarded /the prize /embraced /a friend.
(46) A friend /embraced /the old man /(who was) /awarded /the prize.
(47) The pack leader /(who was) /knitted /a scarf /thanked /the boy scout.
(48) The boy scout / thanked /the pack leader /(who was) /knitted /a scarf.
(49) The player /(who was) /tossed /the ball /interfered with /the other team.
(50) The other team /interfered with /the player /(who was) /tossed /the ball.
(51) The priest /(who was) /mailed /a request /telephoned /the cardinal.[1]
(52) The cardinal / telephoned /the priest /(who was) /mailed /a request.
(53) The principal /(who was) /taught /the policy /emailed /the teacher.
(54) The teacher /emailed /the principal /(who was) /taught /the policy.
(55) The prisoner /(who was) /procured /a weekend pass /whispered to /the guard.
(56) The guard /whispered to /the prisoner /(who was) /procured /a weekend pass.
(57) The reporter /(who was) /told /the rumour /yelled at /the witness.
(58) The witness /yelled at /the reporter /(who was) /told /the rumour.
(59) The scientist /(who was) /shipped /critical supplies /admired /the tribesmen.
(60) The tribesmen /admired /the scientist /(who was) /shipped /critical supplies.
(61) The secretary /(who was) /poured /a beer /chatted up /the lawyer.
(62) The lawyer /chatted up /the secretary /(who was) /poured /a beer.
(63) The settlers /(who were) /built /a log cabin /thanked /the neighbours.
(64) The neighbours /thanked /the settlers /(who were)/ built /a log cabin.
(65) The spy /(who was) /served /a dry martini /watched /the villain.
(66) The villain /watched /the spy /(who was) /served /a dry martini.
(67) The student /(who was) /baked /a cake /called /the roommate.
(68) The roommate /called /the student /(who was) /baked /a cake.
(69) The student /(who was) /called /a bad name /knew /the bus driver.
(70) The bus driver / knew /the student /(who was) /called /a bad name.
(71) The student /(who was) /found /a scholarship /talked to /a professor.
(72) A professor /talked to /the student /(who was) /found /a scholarship.
(73) The teenager /(who was) /burned /a CD /smiled at /the cheerleader.
(74) The cheerleader /smiled at /the teenager /(who was) /burned /a CD.
(75) The traveller /(who was) /booked /the flight /spotted /a cab driver.
(76) A cab driver /spotted /the traveller /(who was) /booked /a flight.
(77) The woman /(who was) /offered /a free drink /winked at /the bartender.
(78) The bartender /winked at /the woman /(who was) /offered /a free drink.
(79) The woman /(who was) /painted /the portrait /waved at /the mayor.
(80) The mayor /waved to /the woman /(who was) /painted /the portrait.

[1]The disambiguated version of this sentence was faulty, so the entire item in both LC and GP conditions was removed from analysis.