Journal of Cognitive Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pecp21

Can propositional biases modulate syntactic repair processes? Insights from preceding comprehension questions

Jack Dempsey & Laurel Brehm

To cite this article: Jack Dempsey & Laurel Brehm (2020) Can propositional biases modulate syntactic repair processes? Insights from preceding comprehension questions, Journal of Cognitive Psychology, 32:5-6, 543-552, DOI: 10.1080/20445911.2020.1803884

To link to this article: https://doi.org/10.1080/20445911.2020.1803884

JOURNAL OF COGNITIVE PSYCHOLOGY
2020, VOL. 32, NOS. 5–6, 543–552
https://doi.org/10.1080/20445911.2020.1803884

Can propositional biases modulate syntactic repair processes? Insights from preceding comprehension questions

Jack Dempsey [a,b] and Laurel Brehm[a,c]

[a]Center for Language Science, The Pennsylvania State University, University Park, PA, USA; [b]Department of Educational Psychology, University of Illinois at Urbana-Champaign, Urbana, IL, USA; [c]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ABSTRACT

There is an ongoing debate about whether discourse biases can constrain sentence processing. Previous work has shown comprehension question accuracy to decrease for temporarily ambiguous sentences preceded by a context biasing towards an initial misinterpretation, suggesting a role of context for modulating comprehension. However, this creates limited modulation of reading times at the disambiguating word, suggesting initial syntactic processing may be unaffected by context [Christianson & Luke, 2011. Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*, *15*(2), 136–166]. The current experiments examine whether propositional and structural content from preceding comprehension questions can cue readers to expect certain structures in temporarily ambiguous garden-path sentences. The central finding is that syntactic repair processes remain unaffected while reading times in other regions are modulated by preceding questions. This suggests that reading strategies can be superficially influenced by preceding comprehension questions without impacting the fidelity of ultimate (mis)representations.

ARTICLE HISTORY
Received 20 March 2020
Accepted 24 July 2020

KEYWORDS
Comprehension questions; temporary ambiguities; syntactic repair

Introduction

Reading can be surprisingly hard, even for skilled readers, and there is variability in how individuals parse hard sentences and in how they recover from errors while reading. One factor known to facilitate processing of difficult sentences is context. For example, Grodner et al. (2005) demonstrated in a self-paced reading experiment that embedding sentences within supportive discourses facilitates their processing. As proposed by Altman and Steedman (1988), this suggests that readers use context to favour a representation that incurs the least processing cost among competitors while reading. In other words, contexts that are biased towards supporting a specific structure within a discourse may facilitate processing.

Temporary ambiguities like in (1c) below are structures that could plausibly benefit from context. These temporary ambiguities lead readers down the "garden path" by biasing the formation of a temporarily valid syntactic structure that is later rendered untenable by the global syntax of the sentence, typically evidenced by longer reading times at the disambiguating word/region (Bever, 1970; Frazier & Fodor, 1978). Earlier studies have shown that readers maintain initial misinterpretations after reading a garden path sentence, as evidenced by higher error rates to questions such as (1d) below compared with questions asking about other parts of the sentence (Christianson et al., 2001, 2006; Qian et al., 2018).

(1a) There was a public outcry against the author of a racy new novel. (S-Bias).
(1b) There was a public outcry against the editor of a racy new novel. (NP-Bias).
(1c) The publisher called up the editor and the author refused to change the book's ending.
(1d) Did the publisher phone the editor and the author?

To investigate whether context modulates the processing difficulty of garden-path structures and deters initial misinterpretations, Christianson and

CONTACT Jack Dempsey jkdemps2@illinois.edu Department of Educational Psychology, University of Illinois, Education Building, Rm. 226A, 1310 S. 6th St., Champaign, IL 61820, USA
Supplemental data for this article can be accessed at https://doi.org/10.1080/20445911.2020.1803884

Luke (2011) conducted three self-paced reading experiments manipulating the propositional content of preceding contexts to bias either a sentential complement interpretation (dispreferred, 1a) or a conjoined NP interpretation (preferred, 1b) of temporarily ambiguous sentences such as (1c). Contexts like (1b) increased error rates to comprehension questions (1d) without impacting reading times at the disambiguating region of the sentence. The authors propose that this implies syntactic reanalysis always occurs in the same way, regardless of any biasing context, and that readers are prone to follow a non-syntactic strategy while reading, preferring instead a "Good-Enough" (incorrect) initial interpretation of a sentence because of a match between the semantic content in the context and the initial parse of the ambiguous item.

Although biases from propositional content have not been shown to facilitate online repair, we suggest that a more direct cue to the desired interpretation of the sentence might be able to do so. Such a cue might bias processing by facilitating syntactic and/or semantic selection processes between the two competing parses (similar to Grodner et al., 2005). In the current experiments, we tested this hypothesis by simply providing the comprehension question to be answered before the sentence was read. The question was: does reversing the order of the materials provide a strong enough bias to influence processing? Presenting a comprehension question that matches the infelicitous interpretation of the sentence lexically, syntactically, and semantically may cue the reader as to what to look out for while reading. In other words, we suggest presenting comprehension questions before temporarily ambiguous sentences could directly cue readers to anticipate a particular structure depending on the wording of the question. We investigated two research questions to assess this hypothesis: (1) Can semantic or syntactic context in the form of transitive-biased preceding questions cause a slowdown in reading times at the disambiguating region of intransitive garden-path sentences and (2) Do these transitive-biased questions lead to higher error rates when presented before the sentences?

Experiment 1

The first experiment used transitive-biased questions about garden-path structures such as in (2a–2d) below. If preceding transitive questions bias readers to expect a transitive structure to a higher degree, we should see evidence of increased difficulty at the spillover word (Mitchell, 1984) for the preceding question conditions. That is: we predicted that preceding questions that biased the (garden path) transitive parse of the sentence, regardless of surface form, should lead to slower reading times at the spillover word due to an increased syntactic bias and lower accuracy for these questions.

We also asked whether the content overlap between the target sentence and the preceding comprehension question would facilitate reading times prior to the disambiguating region. If observed, this could be for several reasons: a speedup in reading time might be due to lexical priming (see Tooley & Traxler, 2010 for a review), syntactic facilitation (as observed in Grodner et al., 2005), or overlapping argument structure (as in Christianson & Luke, 2011). In order to dissociate facilitation due to surface form from facilitation due to argument structure, we manipulated whether the question was presented in the same voice as the sentence (active vs passive). By comparing reading times in early regions of sentences that were preceded by active versus passive questions, we can demonstrate whether priming surface form or argument structure improves processing in the form of faster reading times: if facilitation appears only for active questions, then surface form biases processing, whereas if facilitation appears for both types of preceding questions, argument structure biases processing. Moreover, by comparing reading times at the spillover word, we can assess whether syntactic repair is modulated by structural biases (a slowdown for preceding active relative to passive questions) or propositional biases only (a slowdown for both types of preceding questions).

Methods

Participants

Data were collected from 48 participants recruited on Amazon Mechanical Turk. Qualifications for participation included being a native speaker of English and a United States location.

Materials and design

Critical stimuli were 48 garden-path sentences with Reflexive Absolute Transitive (RAT) verbs (Trask, 1993) like "dressed" and "washed" adapted from Christianson et al. (2001). We manipulated presentation order and the voice of the question to create four versions of each item (as in 2a–2d

below; pre-critical, critical, and spillover words in bold). A 2 × 2 Latin-Square design was used to assign item versions to lists, with an equal number of items in each condition and only one item version presented per list. In addition to the critical items, 64 filler items were constructed with a variety of different structures (e.g. cleft structures); these were also followed or preceded by comprehension questions in both active and passive forms and contained both transitive and intransitive verbs (see Supplementary Materials for full stimuli). In total, all participants therefore read 112 sentence-question pairs. Items were presented in 6 blocks and randomised with filler items within each block.

(2a) While Anna dressed the **baby played in** the crib. (Did Anna dress the baby?)
(2b) While Anna dressed the **baby played in** the crib. (Was the baby dressed by Anna?)
(2c) (Did Anna dress the baby?) While Anna dressed the **baby played in** the crib.
(2d) (Was the baby dressed by Anna?) While Anna dressed the **baby played in** the crib.

Items were presented using the experimental platform Ibex Farm (Drummond et al., 2016). On each trial, participants first saw a fixation cross. They then read items in one of two orders: a question followed by a sentence, or a sentence followed by a question. This was followed by a screen with a single question mark on it, which prompted the participants to answer the question by pressing the F key for "yes" or the J key for "no." All sentences were presented as a series of blanks through which the participant could read at their own pace in a word-by-word non-cumulative fashion by pressing the spacebar. Participants were first given a set of 6 practice items containing variations in sentence-question order and active–passive voice, followed by the 112 experimental trials. After completing the experiment, participants filled out an online questionnaire verifying that they met the qualifications for participation.

Results

Data cleaning

Participants were omitted from analysis if they had less than 80% accuracy for questions to filler items, leading to the exclusion of two participants. Question responses were excluded from all analyses if response times were greater than 11,500 msec, resulting in the omission of eight data points (less than 1%). No question responses were excluded for being fast since the question itself always appeared on a different screen from the answering screen, meaning participants could have been ready to immediately respond. Reading times were excluded if slower than 2000 msec or faster than 100 msec, eliminating less than 1% of data points from analysis.

Reading time analyses

Mean reading times per condition at the pre-critical, critical, and spillover words can be found in the Supplementary Materials. Separate linear mixed effects models containing fixed effects of question order, question voice, word length, and the interaction between question order and question voice were fit for log-transformed reading times at the pre-critical word, critical (disambiguating) word, and the spillover word. Models were first fit with random slopes and intercepts for participant and item. If the model failed to converge, random slopes were eliminated according to the lowest variance contributed. An additional model was run on log-transformed full sentence reading times with the same fixed and random effects structure, except that word length was replaced by sentence length. Question order and voice were sum coded for all mixed effects models (Preceding = −.5, Following = −.5; Active = −.5, Passive = .5). Table 1 shows the results from the three models run using the lme4 package (Bates et al., 2015) in R version 3.5.1 (R Core Team, 2013).

Analyses revealed a main effect of question order at the pre-critical word with preceding questions eliciting faster reading times. Significant effects of question voice and question order were also found at the critical word, indicating significantly faster reading times in conditions with active compared with passive questions and with preceding questions compared with questions following sentences; however, these factors did not interact. No main effects or interactions were found at the spillover word region. Full sentence reading times revealed a main effect of question order with faster reading times after preceding questions, but no significant effect of voice or interaction was found. Figure 1 shows mean reading times per word relative to the critical word across conditions.

Table 1. Exp. 1 RT models.

	Beta	SE	*df*	*t* value	*p* value
Pre-Critical Word					
Voice	0.017	0.013	54.422	1.312	0.195
Presentation	0.069	0.017	54.787	4.076	**<.001**
*Voice*Presentation*	−0.015	0.025	360.462	−0.609	0.543
Critical Word					
Voice	0.054	0.021	353.877	2.558	**0.011***
Presentation	0.067	0.022	55.435	3.027	**.004***
*Voice*Presentation*	−0.031	0.03	353.76	−1.038	0.3
Spillover Word					
Voice	−0.002	0.02	75.338	0.114	0.91
Presentation	0.008	0.027	81.089	0.316	0.753
*Voice*Presentation*	−0.022	0.038	460.924	−0.577	0.564
Full Sentence					
Voice	0.009	0.009	364.523	1.096	0.274
Presentation	0.055	0.014	54.794	3.971	**<.001***
*Voice*Presentation*	−0.005	0.017	364.514	−0.302	0.763

Question accuracy analyses

Means and SDs can be found in the supplemental materials. A logistic mixed effects model with question order, question voice, and their interaction as fixed effects and the maximal random effect structure that converged was fit to the accuracy data with contrasts coded as in the reading time analyses. The model showed no significant main effect of question order or question voice, nor a significant interaction.

Because all the target items in this experiment were temporarily ambiguous, an additional model was fit to all the data (garden path items plus fillers) with item type as a fixed effect to ensure that a classic garden-path accuracy effect occurred (i.e. lower accuracy for garden-path vs filler sentences). There was indeed a significant effect of item type indicating that the error rate for comprehension questions in filler trials was significantly better than for questions in the garden-path items. Results from both models can be found in Table 2.

Discussion

While analyses of question accuracy replicated the classic lingering misinterpretation effect, with higher error rates for questions in garden-path compared to filler trials (Christianson et al., 2001; Qian et al., 2018), we observed no effects of question order or question voice on accuracy. In this experiment, the same patterns of final interpretations were obtained for each condition: it did not matter whether the question preceded or followed the sentence, or whether it syntactically matched or mismatched the sentence.

Reading time analyses revealed a main effect of question order at the critical word, which might appear to suggest that syntactic repair is modulated by context. However, this effect is the opposite direction of what we had predicted: reading times were *faster* for preceding comprehension questions regardless of surface syntax. If semantic content from (garden path) transitive-biased comprehension questions *facilitated* the transitive misparse of the garden-path sentence, the data should have instead shown *slower* reading times at the disambiguating word. Additionally, while ambiguity effects are often found at the spillover word, we observed no effects of voice or question order in that region. Therefore, it is unlikely that the faster reading times arose due to a shift in expectations from the transitive bias of the preceding question.

One possible explanation for the present data is that preceding comprehension questions induced a task-specific strategy in participants, causing them to read the entire sentence faster, which is mirrored by a significant effect of question order in full sentence reading times. A second possible explanation is that lexical priming caused by the overlapping content between question and sentence led to this speedup rather than a global reading strategy. Both are compelling explanations for the present data.

Since all of the target trials in Experiment 1 used transitive-biased comprehension questions, Experiment 1's design does not allow an explicit test of how comprehension changes depending on the preceding question bias. We address this in Experiment 2, where we vary question bias. This also serves to investigate the source of the early facilitation effects observed in Experiment 1.

Experiment 2

In Experiment 2, we sought to replicate the question order effect in Experiment 1 and to test whether

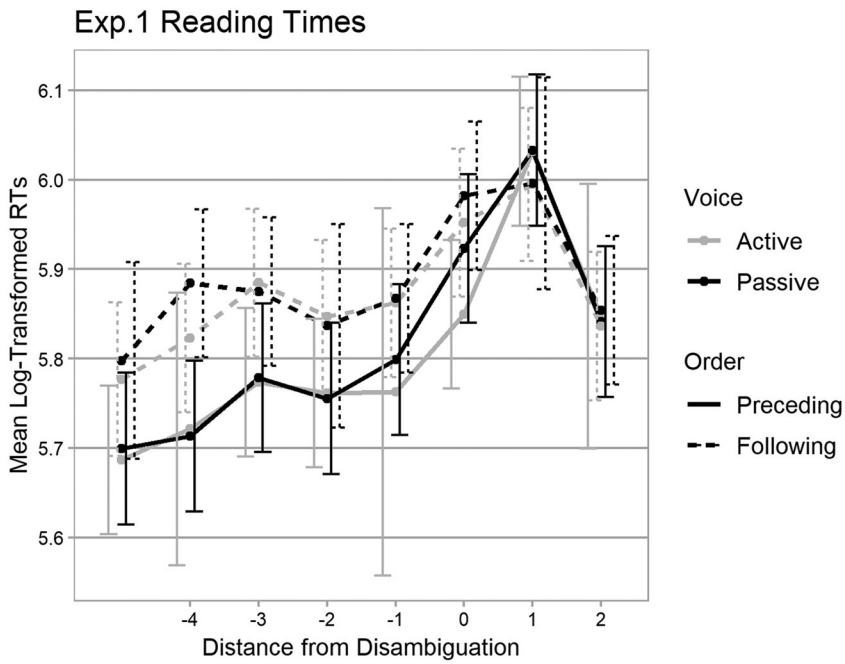

Figure 1. Experiment 1 word-by-word mean log-transformed reading times relative to the critical (disambiguating) word (region 0). Error bars reflect confidence intervals from fitted mixed-effect models.

question biases affect reading times and ultimate comprehension. The goal was to test if propositional content in preceding comprehension questions informs readers' structural expectations and, consequently, their reading patterns and sentence-final interpretations. This would be evidenced by slower reading times at the spillover word and lower accuracy following preceding questions with transitive biases compared to intransitive biases, as the latter may bias the reader toward a correct interpretation of the sentence (i.e. away from the garden path misparse).

In addition, a replication of the speedup in reading times at the pre-critical and critical words across both bias conditions would imply that the preceding question induces a structurally-superficial strategy that causes readers to speed up. Alternatively, if transitive preceding questions, which overlap lexically with the pre-critical word, lead to faster reading times at the pre-critical (and critical) word compared to intransitive questions, this would support a lexical priming account of the question order effect.

Methods

Participants

Data were collected from 48 participants recruited on Amazon Mechanical Turk following the same criteria as Experiment 1.

Materials and design

The 48 critical trials were modified from the previous experiment to have both transitive-biased and intransitive-biased comprehension questions, all in active voice. Examples (3a) and (3c) show the transitive question bias type while (3b) and (3d) show the intransitive question bias type for one item. The same filler questions from Experiment 1 were used for a total of 112 items, and a Latin-Square design was used to assign items in the four versions of question presentation order and bias type into four lists.

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

Results

Data cleaning

Data were cleaned using the same criteria as in the first experiment. No participants were excluded based on question accuracy (all scored greater

Table 2. Exp. 1 accuracy models.

		Beta	SE	*z* value	*p* value
Experiment Model					
	Voice	−0.151	0.188	−0.8	0.424
	Order	0.056	0.244	0.23	0.818
	*Voice*Order*	0.271	0.377	0.718	0.473
Garden-Path Model					
	Condition	−1.607	0.328	−4.896	**<.001***

than 80% on fillers), and less than 1% of the question data and less than 1% reading time data were excluded.

Reading time analyses

Means by condition can be found in the Supplemental Materials. Models as outlined in Experiment 1 were fitted to log-transformed reading times at the pre-critical word, critical word, and spillover word, as well as to log-transformed full sentence reading times with the same fixed and random effects structures as the first experiment, except that bias type replaced question voice. Fixed effects were sum coded as in Experiment 1 (Preceding = −.5, Following = .5; Transitive Bias = −.5, Intransitive Bias = .5), and again, models were first fit with the maximal random effects structure, eliminating random slopes until the model converged. Results are outlined in Table 3.

The models revealed main effects of question order and question bias as well as a significant interaction at the pre-critical word, with faster reading times following preceding questions, especially if the preceding question was biased towards a transitive interpretation. These effects were also found at the critical word. At the spillover word, a significant effect of presentation order was found such that reading times were faster for sentences after preceding questions, and there was a significant main effect of bias type but no reliable interaction.

Full sentence analyses revealed a main effect of presentation order and a non-significant interaction consistent with faster reading of the target sentence following preceding transitive-biased questions. Reading times per word in the sentence are shown in Figure 2.

Question accuracy analyses

Logistic mixed effects models were fit to the data for question accuracy as described in Experiment 1. Transitive-biased questions led to higher error rates than intransitive-biased questions (mirroring the lingering misinterpretation effect), and preceding questions led to higher error rates than questions following the target sentence; however, the two factors did not interact. The results from this model can be found in Table 4.

Discussion

Reading time analyses revealed main effects of question order and bias type and a significant interaction at the pre-critical and critical words of target sentences. As in Experiment 1, preceding questions led to *faster* reading times than questions following the sentence at the pre-critical word and critical word. The main effect of question order coupled with the interaction of question order and bias type at the pre-critical and critical words suggests a role for both a task-induced speedup and lexical priming facilitation. We suggest that the observed patterns are due to a general facilitation at the pre-critical, critical, and spillover words due to task effects, plus an additional effect of lexical priming for the pre-critical and critical words, which further increases reading speed following intransitive-biased questions.

While question bias again influenced ultimate comprehension of the sentence, it did so only as a main effect, replicating the lingering misinterpretation effect seen in Experiment 1 and in earlier work: transitive-biasing questions increased error rates to the same degree whether following or preceding the sentence. However, unlike Experiment 1, we also observed a main effect of question order such that preceding questions generally led to lower accuracy. This effect might be due to the increased variability in the type of comprehension questions, which in Experiment 1 always targeted the early regions of the sentence, but in Experiment 2 targeted either the earlier or later regions of the sentence. This design, together with the working memory load associated with remembering the question while reading the sentence, may have caused more difficulty for participants in Experiment 2. This difference in design could also explain why

Table 3. Exp. 2 RT models.

	Estimate	SE	*df*	*t* value	*p* value
Pre-Critical Word					
Bias	0.027	0.015	42.32	1.883	0.067
Presentation	0.073	0.014	46.67	5.294	**<.001**
*Bias*Presentation*	−0.071	0.02	208.3	−3.534	**<.001**
Critical Word					
Bias	0.033	0.013	46.55	2.52	**.015***
Presentation	0.059	0.018	45.63	3.297	**.002***
*Bias*Presentation*	−0.097	0.025	210.6	−3.893	**<.001***
Spillover Word					
Bias	0.033	0.013	206.6	2.51	**.012***
Presentation	0.04	0.017	47.71	2.271	**.028***
*Bias*Presentation*	0.032	0.026	206.7	1.226	0.22
Full Sentence					
Bias	0.011	0.007	216.4	1.675	0.094
Presentation	0.067	0.01	46.34	6.93	**<.001***
*Bias*Presentation*	−0.023	0.013	217.1	−1.755	0.079

reading was faster at the spillover word in Experiment 2 but not in Experiment 1: it is possible that participants extended their task-based strategy from preceding questions to search the entire sentence for the answer rather than simply the first part of the sentence. Combined, this suggests that reading strategies can be superficially impacted by preceding comprehension questions without influencing the quality of ultimate representations.

General discussion

In two experiments, we showed that, consistent with earlier work (Christianson & Luke, 2011), the parser does not use syntactic or propositional information from preceding context to facilitate the initial disambiguation of temporary ambiguities. This was the case even when the biasing information was the exact comprehension question to be answered later, which suggests direct cuing or priming cannot improve parsing strategies for dealing with ambiguous garden-path structures.

Both experiments also revealed that reading times were superficially modulated by preceding comprehension questions such that overlapping content, in particular, was read more quickly, and preceding questions also led to faster reading overall. This is likely due to a combination of comprehension priming effects (e.g. Tooley & Traxler, 2010), which facilitate words that overlap with the question, and a task-based strategy in which participants selectively choose to read more superficially because preceding questions offer specific content to be either confirmed or disconfirmed in the subsequent sentence.

Although many models of ambiguity resolution take advantage of contextual cues for eventual disambiguation, our study highlights that sentence-level temporary ambiguities lead to some types of processing difficulty that cannot be modulated by the preceding context. This shows how engaging in a transient misparse of a garden-path sentence can lead to lingering misinterpretations, and suggests that preceding context likely does not inform syntactic repair processes. This may be consistent with several processing accounts, as discussed below.

Under a Good-Enough (GE) processing view, difficult or ambiguous structures do not always lead to faithful representations because the goal of the parser is not to exhaustively optimise structure from the input but rather to rapidly build representations that are "good enough" for effective communication (Christianson, 2016; Ferreira et al., 2002; Karimi & Ferreira, 2016). GE processing suggests this occurs because processing happens simultaneously via a slow algorithmic processing stream and a fast heuristic processing stream. Comprehenders engage in shallow processing during reading because the heuristic stream is faster than the algorithmic stream, which is not quick enough to alert the parser to a contradiction between the global syntax and initial misinterpretation. Shallow parsing strategies allow readers to misinterpret garden-path sentences and other types of ambiguities, and in turn make it hard to inhibit these misinterpretations later in processing (evidenced by high error rates) due to the pressure to avoid "disequilibrium" (Karimi & Ferreira, 2016). A strong version of GE processing might also suggest that context, if impacting algorithmic processing, should lead only to late effects, such as lower response accuracy. This pattern was not borne out in the current data: we showed instead that preceding questions affected reading times within the sentence due to

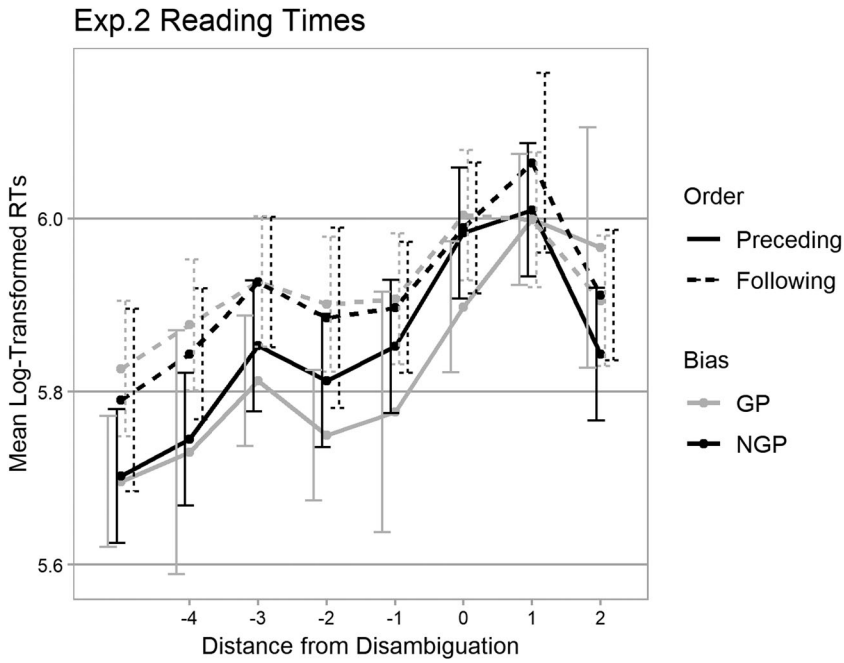

Figure 2. Experiment 2 word-by-word mean log-transformed reading times relative to the critical (disambiguating) word (region 0). Error bars reflect confidence intervals from fitted mixed-effect models.

priming / task-based effects, and sometimes led to decreased response accuracy. This implies that certain types of biases from preceding questions may not directly inform either processing stream. This is contradictory to previous findings suggesting preceding contexts influence offline representations (Christianson & Luke, 2011; Grodner et al., 2005 respectively), and we suggest that the difference is telling about how task instructions influence subsequent reading of text.

An alternative account of lingering misinterpretations is that they arise due to postinterpretive processes (Bader & Meng, 2018) because question accuracy does not veridically measure the quality of sentence-final interpretations but rather reveals artifacts of the memory retrieval process imposed by metalinguistic tasks. While offline tasks play an important role in psycholinguistic inquiry (see Ferreira & Yang, 2019 for an excellent review), it is true that comprehension questions may elicit artifacts derived from recalling linguistic representations. This would suggest that the lack of predicted interaction in the accuracy model is due to the offline nature of the comprehension questions, and the main effect of question order on accuracy in Experiment 2 is attributed to the extra working memory burden for the preceding question condition. Using a task that queries sentence-final representations without increasing working memory burden could provide further evidence for a postinterpretive account where final interpretations *do* take biases from preceding context into account while the online retrieval of these representations is not impacted by the contextual biases.

The current data are therefore consistent with but not adjudicative between these proposed theories. As suggested by either account, we show that misrepresentations happen, may be hard to avoid or recover from, and that at least in some circumstances, readers cannot use contextual information from preceding questions to inform their parsing strategies in real-time.

Importantly, while readers did not benefit from contextual biases, they did seem to benefit from the lexical overlap of the preceding question and the target sentence. Content from preceding comprehension questions that overlapped with the target sentence was read relatively faster at the second presentation. This shows how readers can use at least some information from the context to inform their reading strategies. Why then are readers able to use lexical information but not

Table 4. Exp. 2 accuracy model.

	Beta	SE	z value	*p* value
Bias	3.209	0.279	11.507	**<.001***
Presentation	0.069	0.289	2.389	**0.017**
*Bias*Presentation*	0.593	0.536	1.106	0.269

syntactic or propositional information to update their expectations or facilitate processing of subsequent content? One possibility could be that the transitive structure is already expected, which may be at the heart of why a garden-path effect is observed in the first place. Because the transitive structure is already preferred or expected, a preceding comprehension question may not lead to the updating of any expectations when reading. This would suggest that other types of questions could modulate processing for other types of structures.

This leads to another line of questioning: while our studies suggest participants are not further hindered by preceding biases towards the initial misparse, it remains unanswered whether other types of cue or bias towards the correct parse can ease syntactic repair processes. Preliminary evidence suggests against this. Other work suggests that biases and cuing do not lead to a change in syntactic expectations over the course of an experiment despite evidence for facilitation after repeated exposure to said structure (Harrington-Stack et al., 2018): parsers do not seem to use co-occurring cues in the input (e.g. contextually novel frequency distributions, co-occurring semantic categories or font colour) to avoid or mitigate garden path effects (Dempsey et al., in press). Future work might consider what can and cannot ease syntactic repair, with an eye to how processing facilitation and exacerbation may trade off, or may lead to ceiling or floor effects depending on the process under investigation.

The finding that preceding contexts in the form of questions do not alter question accuracy or syntactic repair in any meaningful way is also informative when contrasted with the role of contextual biases from discourse. The current work differs from the existing discourse literature in that preceding questions may not be part of the discourse context per se, but instead explicitly cue readers towards upcoming input. This makes them more like task instructions (i.e. "find the right answer"), as we have suggested above. Studies investigating the role of task instructions on reading suggest the type of instruction given to participants can influence how they allocate attention during reading (Kaakinen & Hyönä, 2010; Schotter et al., 2014), and work in sentence processing has shown that the mere presence of comprehension questions alters reading patterns by invoking pressure on participants to attempt to more deeply process the material (Dwivedi, 2013; Swets et al., 2008). Our findings add to this growing body of evidence that greater allocation of attention towards text, either as a function of text-based interest (McDaniel et al., 2000) or as a function of time spent reading/rereading (Christianson & Luke, 2011; Christianson, 2016), does not necessarily inform successful repair or more accurate recall. This in turn suggests that allocation of attention during reading may play a minimal role in resolving temporary ambiguities. Further work is needed to better dissociate the effects of task-instructions from discourse contexts in experimental settings.

In conclusion, temporary ambiguities cause processing difficulty that is likely not modulated by preceding cues. Our results suggest that effects of ambiguity linger irrespective of strong contextual cues and thereby demonstrate the importance of understanding how ambiguities can affect processing further downstream.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data from these experiments are accessible online. doi:10.17605/OSF.IO/82V4Q.

ORCID

Jack Dempsey http://orcid.org/0000-0003-1267-6780

References

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238. https://doi.org/10.1016/0010-0277(88)90020-0

Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1286–1311. https://doi.org/10.1037/xlm0000519

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bever, T. G. (1970). The cognitive basis for linguistic structures. *Cognition and the Development of Language*, *279* (362), 1–61.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good- enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 817–828. https://doi.org/10.1080/17470218.2015.1134603

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Christianson, K., & Luke, S. G. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*, *15*(2), 136–166. https://doi.org/10.1080/10888431003636787

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*(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6

Dempsey, J., Liu, Q., & Christianson, K. (in press). Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Drummond, A., von der Malsburg, T., Erlewine, M. Y., & Vafaie, M. (2016). *Ibex Farm*. GitHub. https://github.com/addrummond/ibex

Dwivedi, V. D. (2013). Interpreting quantifier scope ambiguity: Evidence of heuristic first, algorithmic second processing. *PloS One*, *8*, 11. https://doi.org/10.1371/journal.pone.0081461.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15.

Ferreira, F., & Yang, Z. (2019). The problem of comprehension in psycholinguistics. *Discourse Processes*, *56*(7), 485–495. https://doi.org/10.1080/0163853X.2019.1591885

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Grodner, D., Gibson, E., & Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition*, *95*(3), 275–296. https://doi.org/10.1016/j.cognition.2004.01.007

Harrington Stack, C. M., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition*, *46*(6), 864–877. https://doi.org/10.3758/s13421-018-0808-6

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1561–1566. https://doi.org/10.1037/a0020693

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

McDaniel, M. A., Waddill, P. J., Finstad, K., & Bourg, T. (2000). The effects of text-based interest on attention and recall. *Journal of Educational Psychology*, *92*(3), 492. https://doi.org/10.1037/0022-0663.92.3.492

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 69–89). Erlbaum.

Qian, Z., Garnsey, S., & Christianson, K. (2018). A comparison of online and offline measures of good-enough processing in garden-path sentences. *Language, Cognition and Neuroscience*, *33*(2), 227–254. https://doi.org/10.1080/23273798.2017.1379606

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for statistical Computing.

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition*, *131*(1), 1–27. https://doi.org/10.1016/j.cognition.2013.11.018

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216. https://doi.org/10.3758/MC.36.1.201

Tooley, K. M., & Traxler, M. J. (2010). Syntactic priming effects in comprehension: A critical review. *Language and Linguistics Compass*, *4*(10), 925–937. https://doi.org/10.1111/j.1749-818X.2010.00249.x

Trask, R. L. (1993). *A dictionary of grammatical terms in linguistics*. Routledge.