Discourse coherence modulates use of predictive processing during sentence comprehension

Citation for published version:
Carter, G & Hoffman, P 2024, 'Discourse coherence modulates use of predictive processing during sentence comprehension', *Cognition*, vol. 242, 105637. https://doi.org/10.1016/j.cognition.2023.105637

Digital Object Identifier (DOI):
10.1016/j.cognition.2023.105637

Published In:
Cognition

Cognition 242 (2024) 105637

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Discourse coherence modulates use of predictive processing during sentence comprehension

Georgia-Ann Carter [a,c,*], Paul Hoffman [b,c]

[a] Institute for Language, Cognition and Computation, School of Informatics
[b] School of Philosophy, Psychology and Language Sciences
[c] University of Edinburgh, UK

ARTICLE INFO

Keywords:
Self-paced reading
Discourse coherence
Predictive processing

ABSTRACT

Context has been shown to be vitally important for comprehension. Lexical processing is facilitated when words are highly predictable given their local sentence context, suggesting that people pre-activate likely upcoming words to aid comprehension. However, this facilitation is affected by knowledge about the global context in which comprehension takes place: people predict less when in an environment where expectations are frequently violated. The current study investigated whether discourse coherence is an additional cue that comprehenders use to modulate lexical prediction. In a series of online, self-paced reading experiments, participants read target sentences preceded by short contextual preambles. Local facilitation effects were manipulated through the cloze probability of a critical word within the target sentence and discourse coherence was manipulated by varying the degree to which the target sentence was consistent with the information presented in the preamble. In the first two experiments, target sentences were read more slowly when they occurred in less coherent discourses, but no local facilitation effects were observed. In the third experiment, we strengthened the predictability manipulation by using semantically anomalous critical words. In this experiment, predictable words were processed more quickly and anomalous words more slowly when they occurred in highly coherent discourse. Our results suggest that comprehenders are sensitive to shifts in the topic of discourse and that they downregulate predictive processing when they encounter incoherence in the discourse. This is consistent with recent theoretical accounts suggesting that comprehenders flexibly engage in predictive processing, pre-activating semantic and lexical information less when their expectations are less likely to be reliable.

1. Introduction

Local context has a rapid influence on lexical processing. For example, there is a wide and established literature on the reduction in N400 amplitudes for target words that are semantically related to previous linguistic information (see Kutas & Federmeier, 2011; Kutas & Hillyard, 1980; Kutas & Hillyard, 1984; Van Petten & Luka, 2012). When words are predictable based on their local sentence context, they are processed more quickly, which suggests that the processing mechanisms involved are anticipatory (Schustack, Ehrlich, & Rayner, 1987). While much research has focused on understanding the nature of this anticipatory process, the extent of its influence and the types of information that contribute to it remain under debate (Huettig, 2015; Van Petten & Luka, 2012). A range of linguistic properties have been implicated in these contextual influences, including the semantic properties of the preceding linguistic input (Kutas & Hillyard, 1984), the grammatical gender of a referent (Fleur, Flecken, Rommers, & Nieuwland, 2020; Nieuwland et al., 2018; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2004), and the definiteness of a referent (Burkhardt, 2006; Carter & Nieuwland, 2022; Fleur, Flecken, Rommers, & Nieuwland, 2020). As such, it appears that cues at various levels of the language system are used during comprehension to facilitate the processing of upcoming information (Morris, 2006; Rabovsky & McClelland, 2020; Willems & Peelen, 2021). In the present study, we investigate the effect of cues at the lexical level (Experiments 1 and 2) and the semantic level (Experiment 3).

However, for comprehension to occur, information from the current sentence has to be integrated with prior information from the

* Corresponding author at: Institute for Language, Cognition and Computation, School of Informatics, University of Edinburgh, Informatics Forum, 10 Crichton Street, Newington, Edinburgh EH8 9AB, UK.

E-mail address: georgia.carter@ed.ac.uk (G.-A. Carter).

https://doi.org/10.1016/j.cognition.2023.105637
Received 1 July 2022; Received in revised form 13 April 2023; Accepted 5 October 2023
Available online 17 October 2023

surrounding discourse context (Myers & O'Brien, 1998). To understand the discourse, people form a mental/situation model of it (Kintsch, 1998; van Dijk, 2006; Zwaan, 2016; Zwaan & Radvansky, 1998). We predicted that the status of this discourse model influences the processing mechanisms comprehenders engage in when dealing with sentence-level constraints. As such, the present study investigates how the use of local within-sentence constraints interacts with the properties of the broader discourse in which the sentence occurs.

1.1. Evidence for discourse facilitation effects

Individual sentences are typically preceded by a broader discourse context, which is in itself a rich source of semantic and conceptual information. The presence of a global discourse context facilitates processing for globally related concepts (Albrecht & O'Brien, 1993; Camblin, Gordon, & Swaab, 2007; van Berkum, Hagoort, & Brown, 1999), through the formation of a mental model of the characters and events involved (van Dijk, 2006). A number of models of this process have been proposed. Kintsch's construction-integration model places a proposition as the key component, suggesting that a mental representation of a text is a network of these propositions, linked by shared arguments (Kintsch, 1998). Zwaan and colleagues have proposed a situation model account, whereby an event-based mental representation of the current state of affairs is formed (Zwaan, 2016; Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998). They propose that comprehenders update these representations autonomously during integration of the unfolding linguistic input. Depending on the overlap of semantic features from the incoming input and the current situation model, the current model is either maintained (i.e., high featural overlap), or updated to take new information into account (i.e., low featural overlap). It has been suggested that this is an automatic and continuous pattern-matching process (Myers & O'Brien, 1998). In the present study, we are interested in how this process of maintaining a model of the entire discourse interacts with the use of constraints within a sentence to facilitate comprehension. Throughout the study, we define the local context as the sentence currently being read, whilst the global context refers to the higher-level mental model that is maintained across multiple sentences (Myers, O'Brien, Albrecht, & Mason, 1994; Zwaan, 2016).

So how does the wider global discourse impact the anticipatory processing advantages that are observed within local sentences? There is mixed evidence on this issue (Ledoux, Camblin, Swaab, & Gordon, 2006). In one important study, Hess, Foss, and Carroll (1995) attempted to distinguish between lexical explanations and discourse-based explanations for context effects using a naming task. Participants heard short passages, saw the final word of the passage (the target word) on-screen and were asked to name this. The researchers manipulated how well the target word related to the global context and the local sentence, separately. They found that, without the addition of the prior context, naming latencies were facilitated if associatively related words were present in the sentence (e.g., "The *English/computer science* major wrote the **poem**"). This confirmed the presence of local facilitation effects. Subsequent experiments added global discourse contexts that were either related or unrelated to the final sentence. Naming latencies in the global related conditions were consistently faster than those in the global unrelated conditions. However, local facilitation effects were no longer reliably observed. In other words, a local facilitation effect was evident when the sentences were presented in isolation, but was not visible when a supportive global discourse context was added in subsequent experiments. The authors concluded that discourse context plays a critical role in lexical processing and can override the effects of local constraints (for related findings, see Albrecht & O'Brien, 1993). However, one potential issue in this study is that the contexts used for the global unrelated conditions were not always the same across the local manipulation, which may have led to a confound.

Other studies have provided different perspectives on the roles of global and local context in lexical facilitation. For example, Traxler, Foss, Seely, Kaup, and Morris (2000) manipulated both the association and plausibility of a target word, given the context (e.g.., "The *lumberjack/young man* carried/chopped the **axe** early in the morning"). In this example, "axe" is the target word, which is plausible in its immediate local context when preceded by the verb "carried", and is associated with the wider context when preceded by "lumberjack". Reading time measures were generally influenced by the plausibility of the immediate context but not the wider context, suggesting that global context does not override local plausibility constraints. Conversely, another study investigating how comprehenders combine information from both the local sentence and global contexts found that discourse-level information did impact the processing of upcoming words (Schwanenflugel & White, 1991). These authors conducted three experiments which varied in task (lexical decision vs. naming) and the distance of global information to the local target sentence (far vs. near). They concluded that discourse information interacts with information from the local context, with both sources of constraint benefiting the processing of upcoming words. These studies indicate that there is inconsistency in the literature on the relationship between global and local contexts, and their subsequent influence on processing.

Further evidence of the impact of global context on semantic facilitation effects has been found with neuroimaging. Camblin et al. (2007) used both eye-tracking and ERPs to investigate how discourse-level representations affect within-sentence priming. The authors manipulated semantic association between a local prime and the sentence-final target word, as well as congruency between prior discourse context and the target word. Their first experiment demonstrated a reduction in N400 amplitudes for congruent conditions and associated conditions, while the second experiment also found shorter regression-path reading times for congruent and associated conditions. These findings suggest that both local and global constraints influence lexical processing. In subsequent experiments, the authors disrupted the presence of a coherent discourse by either presenting target sentences with no prior context or with an incoherent context made up of unrelated sentences. Under these circumstances, local association effects were stronger than when sentences were presented within a coherent discourse. From this, they concluded that discourse congruence had a robust and rapid effect on both ERPs and eye-tracking, whereby the facilitation from local lexical associations was strongest when there was no coherent discourse information to support processing (Camblin et al., 2007).

The possibility of an interaction between local and global contextual cues has been further explored by Boudewyn, Long, and Swaab (2015). These authors conducted an ERP study using auditory stimuli to investigate how cues from both global and local contexts contributed to two levels of prediction (semantic features vs. lexical form). The authors created two-sentence mini-stories, manipulating global predictability of the critical noun given the discourse context and local consistency of the critical noun compared with an associated feature word within the local sentence (e.g., "Frank was throwing a birthday party, and he had made the dessert from scratch. After everyone sang, he sliced up some sweet/healthy and tasty cake/veggies that looked delicious"). The authors found a graded N400 response, whereby N400 amplitudes were influenced by both local and global manipulations. Moreover, evidence of an interaction suggested that support from one level was able to compensate for ambiguity and inconsistency at the other.

1.2. Mechanisms supporting contextual facilitation

The studies discussed above provide a mixed view of how global and local cues interact to facilitate processing. Why do the effects of the global discourse seem to be so variable? One possibility is that use of predictive processing itself is flexible and varies depending on its likely utility. In other words, people may use cues from the wider experimental context to determine how valid their expectations are likely to be, and thereby modulate how strongly to engage in predictive processing. A number of studies have shown that extralinguistic cues can affect the degree to which local predictability influences processing, supporting the idea that predictive processing is dynamic and flexible (Brothers, Dave, Hoversten, Traxler, & Swaab, 2019; Brothers, Swaab, & Traxler, 2017; Hagoort, 2004; Hald, Steenbeek-Planting, & Hagoort, 2007). For example, in an ERP study, Brothers et al. (2019) manipulated speaker reliability. Participants heard sentences read by different speakers. Critical sentences were identical for both speakers. However, high-cloze filler sentences were used to manipulate the characteristics of each speaker, such that the reliable speaker usually produced predicted sentence completions, while the unreliable speaker frequently produced completions that violated the listener's predictions. Brothers et al. (2019) argued that a purely automatic prediction mechanism would not lead to differences in N400 amplitudes across the manipulation of speaker reliability, as the linguistic context remained the same. In contrast, a flexible account would predict that comprehenders would be sensitive to the build-up of knowledge about the speaker across the experiment and subsequently use that to tailor their expectations. Supporting the latter account, they found an interaction between speaker reliability and lexical predictability on N400 amplitudes, such that larger effects of contextual constraint were observed when the speaker was reliable, compared to unreliable. This suggests that language comprehension recruits a type of dynamic flexible expectancy mechanism, which is sensitive to extralinguistic cues about the likely value of context-based predictions. When participants listened to a speaker whose speech regularly defied expectations, they became less likely to predict words in advance.

This type of flexibility was also investigated in an earlier study by Brothers et al. (2017), which focused on whether top-down goals and strategies influenced predictive processing across both sentences and wider discourses. The findings provide further evidence that extralinguistic context modulates predictive processes. In their first experiment, larger N400 modulations were found when participants were explicitly instructed to predict the upcoming target word, rather than just passively comprehend the stimuli. Moreover, in their second experiment, the authors found a smaller effect of predictability on reading times when the overall experimental environment was less supportive of predictive processing. The experimental environment was manipulated by changing the proportion of filler sentences that ended in highly predictable or less predictable endings. The assumption was that if comprehenders dynamically modulate their use of predictive processing, the effect of predictability should be reduced when the experiment contained a greater proportion of less predictable sentences. This is exactly what was found. As such, this finding provides further evidence for the influence of extra-linguistic cues on the degree to which context facilitates processing. It shows that local facilitation effects are attenuated when participants do not expect the sentence they are reading to provide reliable cues to its completion. In the present study, we investigated whether a similar attenuation effect would occur when readers encounter a shift in the global discourse topic.

1.3. The current study

In sum, there is a range of evidence with mixed results as to the interaction of global and local contextual cues on discourse comprehension, both in behavioural and neuroimaging studies (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Just & Carpenter, 1980; Just, Carpenter, & Woolley, 1982; McDonald & Shillcock, 2003; Traxler & Foss, 2000). One possible explanation for the variety of findings is that the use of predictive processing is flexible and strategic, such that comprehenders are sensitive to cues about the reliability of predictive processing and use this to regulate the degree to which they use local context to facilitate word recognition. In the present study, we investigated this by manipulating the coherence of discourse in three-sentence written passages (see Table 1). Coherence simply refers to the degree to which discourse forms a series of well-connected statements that can be considered a meaningful whole (Ellis, Henderson, Wright, & Rogalski, 2016). When discourse lacks coherence (i.e., when we encounter unexpected or contradictory information within the discourse), we need to reconfigure the active situation model to accommodate this (Kintsch, 1998; Kuperberg, 2021; Zwaan, 2016; Zwaan & Radvansky, 1998). A lack of coherence could also indicate greater uncertainty about the reliability of local processing cues. Thus, if predictive processing as a mechanism is flexible and dynamic, people may use a lack of coherence in discourse as a cue to temporarily reduce the influence of local context on processing until a stable situation model has been re-established. In contrast, if context facilitates comprehension in an automatic fashion, local context effects should not be sensitive to disruption at the discourse level.

Table 1
Example sentences across four conditions in Experiments 1 and 2.

Condition	Context sentences	Target sentence
High Coherence, High Predictability	Crime was a problem. Residents were starting to feel unsafe.	The police arrested the local gang for selling **drugs** on the corner.
High Coherence, Low Predictability		The police arrested the local gang for selling **toys** on the corner.
Low Coherence, High Predictability	The city was a safe place. It had a low crime rate and high employment.	The police arrested the local gang for selling **drugs** on the corner.
Low Coherence, Low Predictability		The police arrested the local gang for selling **toys** on the corner.

Note. Critical word in target sentence presented in bold.

To this end, we conducted three online, self-paced reading studies where we manipulated both the predictability of the critical word within the target sentence (as a local constraint), and the coherence of the target sentence with the preceding discourse (as a global constraint). This enabled us to investigate whether local facilitation effects are suppressed when comprehenders process less coherent passages.

We recognise that anticipation of information across multiple linguistic levels (syntactic, lexical, semantic, and possibly phonological) can contribute to facilitation for expected continuations in sentences (Heilbron, Armeni, Schoffelen, Hagoort, & de Lange, 2022; Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Kuperberg, 2021; Kuperberg, Brothers, & Wlotko, 2020). Accordingly, we manipulated whether stimuli violated expectations at the lexical level (Experiments 1 and 2) and at the semantic level (Experiment 3). Violation of expectations at the lexical and semantic levels have similar effects on markers of predictive processing like the N400, though effects of semantic violations are often observed later in processing (DeLong, Quante, & Kutas, 2014; Haeuser & Kray, 2022; Nieuwland et al., 2020; Quante, Bölte, & Zwitserlood, 2018).

2. Experiment 1

2.1. Methods

Methods and analysis strategies for this experiment were pre-registered at https://osf.io/h73qg. Where procedures deviated from the pre-registered plan, we have noted this in the text. Further, all data associated with the current study is available on OSF at https://osf.io/a8j4c/.

2.1.1. Participants

Ethical approval for all studies was granted by the School of Philosophy, Psychology & Language Sciences Research Ethics Committee (ref: 241-1920/4) at the University of Edinburgh and informed consent was obtained for all participants. Participants were recruited either on the University of Edinburgh's SONA platform or the Testable Minds online participant pool. Participants recruited through SONA were firstyear Psychology undergraduates who were native speakers of English and were reimbursed with course credits. Participants recruited through Testable Minds consisted of a range of backgrounds; however, we required that the participants were based in the UK, had an approval rating >90%, were native speakers of English and had a maximum age of 40. (Note: Testable Minds defines the approval rating as "the ratio of the number of approved results of a participant to the total number of completed studies"). In this way, we tried to minimise the demographic differences between each participant group. The use of an online study actually allowed us to reach a more diverse participant pool than an undergraduate sample (Enochson & Culbertson, 2015). Previous evidence from both reaction time experiments and comparison reviews have demonstrated that online experiments are a viable method for data collection, including for self-paced reading paradigms (Anwyl-Irvine, Dalmaijer, Hodges, & Evershed, 2021; Enochson & Culbertson, 2015; Johnson, Dayan, Censor, & Cohen, 2022). Participants from Testable Minds were reimbursed with $5.60.

Prior to data collection, we performed a power calculation to determine the approximate sample size required. Because we had no prior estimate of the effect size for the interaction between coherence and predictability, we utilised a conservative effect size estimate of Dz = 0.4, an alpha of 0.05 and power of 0.80. This returned an estimated sample size of 52 participants. We used a two-step process of data collection whereby we pre-registered our intention to initially collect 52 participants' worth of data, check to see if there was a significant interaction between coherence and predictability, and if not, to collect another batch of 52. If there was already a significant interaction present, then we would have stopped data collection (Pocock, 1977). We used this approach to ensure that our study would be well-powered, given that we did not have strong a priori expectations of the effect size. For our pre-registered analyses, we adjusted our alpha from 0.05 to 0.029 to account for this two-stage approach (Pocock, 1977). We report the results from our pre-registered analyses on our initial batch of participants in the supplementary materials (see Fig. A2).

In total, we collected data from 134 participants who indicated having English as their first language, with 80 from the pool of Psychology students and 54 from Testable Minds (mean age: 23, range: 18–42, SD = 6.77). Our final sample size consisted of 84 participants, as 50 participants were excluded for the reasons stated below (see Analysis – Pre-registered).

2.1.2. Design

Participants read passages consisting of a two-sentence preamble and a target sentence. The study used a 2 × 2 within-participants manipulation of passage coherence and critical word predictability. The predictability of a critical word in each target sentence was manipulated using cloze probabilities. The coherence of passages was manipulated by varying the degree to which the target sentence was consistent with the context established by the preamble.

2.1.3. Stimuli

The stimulus set was developed from a prior dataset of sentences that included cloze completion values (Peelle et al., 2020). We filtered the sentences with a cloze probability of >0.75 and selected a batch of 200 candidate sentences which became our target sentences. To manipulate predictability, we swapped the high cloze completion word for a less predictable continuation. We ensured that the less predictable word had the same part-of-speech and existed within the same semantic domain as the highly predictable completion (see Table 1). We calculated the cloze probabilities for these less predictable words using the original Peelle et al. (2020) dataset. Where none of the participants in Peelle et al. provided a particular completion, we assigned it a cloze probability of 0. We added three "spillover" words to the sentences, after the critical word, in order to account for the nature of self-paced reading time effects, which are typically extended over time (Jegerski, 2013).

To manipulate coherence, we created high and low coherent two-sentence context preambles. For each target sentence, we began by creating a highly coherent preamble which provided a strong and highly meaningful context in which to process the target sentence. We then generated a low coherence preamble for each target sentence. The low coherence preambles were designed to be in the same general semantic domain as the target sentence to ensure they could be interpreted as part of a single discourse and such that the target sentence would either depart from the topic established in the preamble or would contain new or contradictory information. The high and low coherence versions of the preambles were constructed with a similar syntactic structure to ensure the stimuli remained as naturalistic as possible, and to avoid a change in processing demands (e.g., processing differences between active vs. passive constructions; Olson & Filby, 1972). As such, we had sentences spread across four conditions, with an example trial presented in Table 1.

2.1.4. Norming

120 participants rated the coherence of the preambles with the target sentences on a 5-point Likert scale (1: Not Coherent to 5: Very Coherent). They were presented with each preamble plus target as a single paragraph and were advised that coherence refers to the degree to which the sentences in the paragraph are meaningfully connected with each other and make sense as a whole, and to rate the passages accordingly. Each participant only saw one version of each trial (HCHP, HCLP, LCHP or LCLP). Participants were recruited from both the University of Edinburgh SONA recruitment platform and Amazon Mechanical Turk (AMT). 62 participants were rejected for failing attentional checks, thus in total, 58 surveys were analysed.

From analysis of the ratings for the 200 normed sentences, the best 160 sentences were selected whose HC and LC coherence ratings showed the largest difference, while controlling for other relevant psycholinguistic properties reported in Table 2. As such, we had a final set of 160 stimuli, with four versions of each passage across our four conditions. Properties of the stimuli in each condition are shown in Table 2. We ensured that HC and LC preambles were matched for length (number of words; $t = -1.11$, $p = 0.27$) and that HP and LP critical words were matched for frequency ($t = 0.98$, $p = 0.33$) and bigram frequency ($t = 1.03$, $p = 0.31$). However, the length of the critical words did differ slightly between the HP and LP conditions ($t = -2.29$, $p = 0.02$). We accounted for this within our pre-registered linear mixed-effects models for Experiment 1 by including critical word length as a covariate. For all other analyses, critical word length was included in our initial linear mixed-effects model as a fixed effect. We derived word frequencies from SUBTLEX-UK (van Heuven, Mandera, Keuleers, & Brysbaert, 2014) and bigram frequencies from MCWord (Medler & Binder, 2005). Differences in coherence ratings between conditions was assessed using a 2 × 2 within-items ANOVA. As expected, HC trials were rated as more coherent than LC trials ($F = 988.08$, $p = 1.36 \times 10^{-131}$). HP trials were also slightly more coherent than LP trials ($F = 18.87$, $p = 1.63 \times 10^{-5}$), presumably because the presence of a less predictable critical word

Table 2
Average psycholinguistic properties of stimuli for Experiments 1 and 2.

Property	HC/HP	HC/LP	LC/HP	LC/LP
Coherence	3.83 (0.46)	3.57 (0.60)	2.46 (0.52)	2.36 (0.48)
Preamble Length	13.54 (2.05)	13.54 (2.05)	13.78 (1.92)	13.78 (1.92)
Critical Word Length	4.35 (1.16)	4.56 (1.09)	4.35 (1.16)	4.56 (1.09)
Critical Word Zipf Frequency	4.53 (0.80)	4.45 (0.88)	4.53 (0.80)	4.45 (0.88)
Critical Word Bigram Frequency	1794.57 (1536.06)	1639.93 (1208.17)	1794.57 (1536.06)	1639.93 (1208.17)
Cloze Value	86.23 (6.59)	0.24 (1.17)	86.23 (6.59)	0.24 (1.17)

Note. Standard deviations are shown in parentheses. HC = high coherence; HP = high predictability; LC = low coherence; LP = low predictability.

affected raters' perceptions of the coherence of the whole passage. However, there was no significant interaction between factors ($F = 3.31$, $p = 0.07$). To ensure that coherence differences between HP and LP could not account for the experimental effects, we modelled coherence using trial-specific coherence values in our analyses (in exploratory analyses for Experiment 1 and pre-registered analyses for Experiments 2 and 3).

2.1.5. Procedure

The study was coded in jsPsych (de Leeuw, 2015). Four lists, counterbalancing the assignment of each sentence to the four conditions, were created, alongside four additional lists with the trial order reversed to avoid order effects. Thus, each participant saw one of the four versions of each passage. The 160 trials were split into four blocks, with a break between each block, the length of which was controlled by the participants. On each trial, the two-sentence preamble was first displayed on-screen in full. After reading this, participants pressed the space bar to trigger the target sentence. The target sentence was shown one word at a time, with each word centred on the screen. Participants moved to the next word in the target sentence by pressing the spacebar. Intertrial intervals were marked with a cross shown on-screen for 750 ms.

To encourage participants to attend to the materials, yes/no comprehension questions were presented immediately after 20% of the target sentences. These were spread throughout the experiment and were evenly distributed across conditions. Answers were balanced for requiring information from either the preambles or the target sentences. Feedback was provided after each question to encourage participants to pay more attention if needed. Sentences were presented in black, 24-point Open Sans font on a white background. Comprehension question feedback was presented in coloured font depending on the outcome (green for correct and red for incorrect).

2.1.6. Analysis

2.1.6.1. Pre-registered. Prior to data collection, we submitted a pre-registration of our intended analyses. We divided each trial into three regions of interest for analysis: 1) preambles (two-sentence contexts), 2) pre-critical word (from the beginning of the target sentence up to, but not including, the critical word) and 3) critical word plus spillover (see Fig. 1). Our main interest was in the processing effects on the critical word. However, in self-paced reading paradigms, behavioural effects are typically extended over time beyond the critical word itself (Jegerski, 2013). Therefore, it is the third ROI (critical+spillover) which is most informative for our hypotheses. However, for completeness, we also included a supplementary analysis of reading times for the critical word alone, without the spillover region. Our dependent variable was the mean reading time of the words within the ROI. Each ROI underwent winsorisation of RTs occurring outside the range of 2 SDs from the mean, calculated for each participant. We also performed log transformations on RTs for the target sentence ROIs.

We fit linear mixed effects models in R for each ROI using the lme4 package (Bates, Mächler, Bolker, & Walker, 2014; R Core Team, 2018). We included fixed effects of Coherence, Predictability and their interaction, as well as specifying length and Zipfian frequency of the critical word as covariates in ROIs that included the critical word. For our categorical predictors, we used sum contrast coding. We also included by-participant and by-item random intercepts. For model fitting, we used a maximal approach, whereby if the maximal model did not converge, we altered the model in a step-wise fashion until convergence (Barr, Levy, Scheepers, & Tily, 2013). These steps were, in order, removal of random correlations, removal of random slopes for the interaction terms and removal of random slopes for the main effects. Summary statistics were computed using lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017), which provides *p*-values using Satterthwaite's degrees of freedom method. For all models, we have given the pruned LMM equation in the Supplementary Materials.

We also pre-registered exclusion criteria, which consisted of excluding a participant if their comprehension question accuracy was lower than 80% ($n = 50$). Therefore, the final sample consisted of 84 participants. Additionally, we excluded trials that prompted an incorrect answer to a comprehension question ($n = 327$).

2.1.6.2. Exploratory analyses. After collecting the data and performing the pre-registered analyses, we re-analysed the data with an improved model, as described in the Results. As these changes were not pre-registered, we label these as exploratory for Experiment 1.

2.2. Results

2.2.1. Pre-registered analyses

Following our pre-registered analysis plan, we ran four linear mixed effects models, each assessing the effects of coherence of the target sentence with the preamble, predictability of the critical word and their interaction. Estimated means for each ROI are presented in Fig. 2.

2.2.1.1. Preambles. Participants had similar reading times for high and less coherent preamble contexts ($\beta = -0.001$, $SE = 0.009$, $t = -0.14$, $p = 0.89$), as well as for high and less predictable critical words ($\beta = 0.004$, $SE = 0.006$, $t = 0.66$, $p = 0.51$), with no interaction ($\beta = -0.002$, $SE = 0.006$, $t = -0.29$, $p = 0.77$). These results confirm that participants took a similar time to read the preambles across all conditions.

2.2.1.2. Pre-critical word. For the pre-critical ROI, we found that processing was facilitated (i.e., faster reading times) when preceded by highly coherent preambles ($\beta = -0.006$, $SE = 0.002$, $t = -3.07$, $p = 0.002$). In contrast, there was no difference in reading times between high and low predictable conditions ($\beta = 0.0007$, $SE = 0.002$, $t = 0.40$, $p = 0.69$), which was expected as this region precedes the critical word. Further, no interaction effect was observed ($\beta = -0.002$, $SE = 0.002$, $t = -1.10$, $p = 0.27$).

2.2.1.3. Critical word and spillover. At the critical word and spillover region, our main ROI, there was no processing advantage for critical words preceded by high coherent conditions, compared to less coherent ($\beta = -0.004$, $SE = 0.002$, $t = -1.56$, $p = 0.12$), nor for high predictable critical words compared to less predictable ($\beta = -0.005$, $SE = 0.003$, $t =$

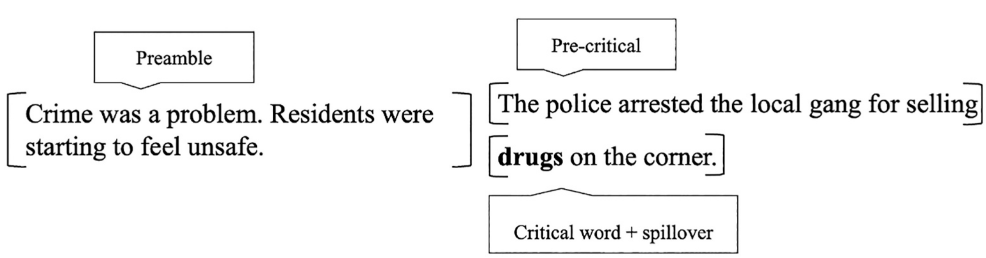

Fig. 1. Example trial separated into ROIs used during analysis.

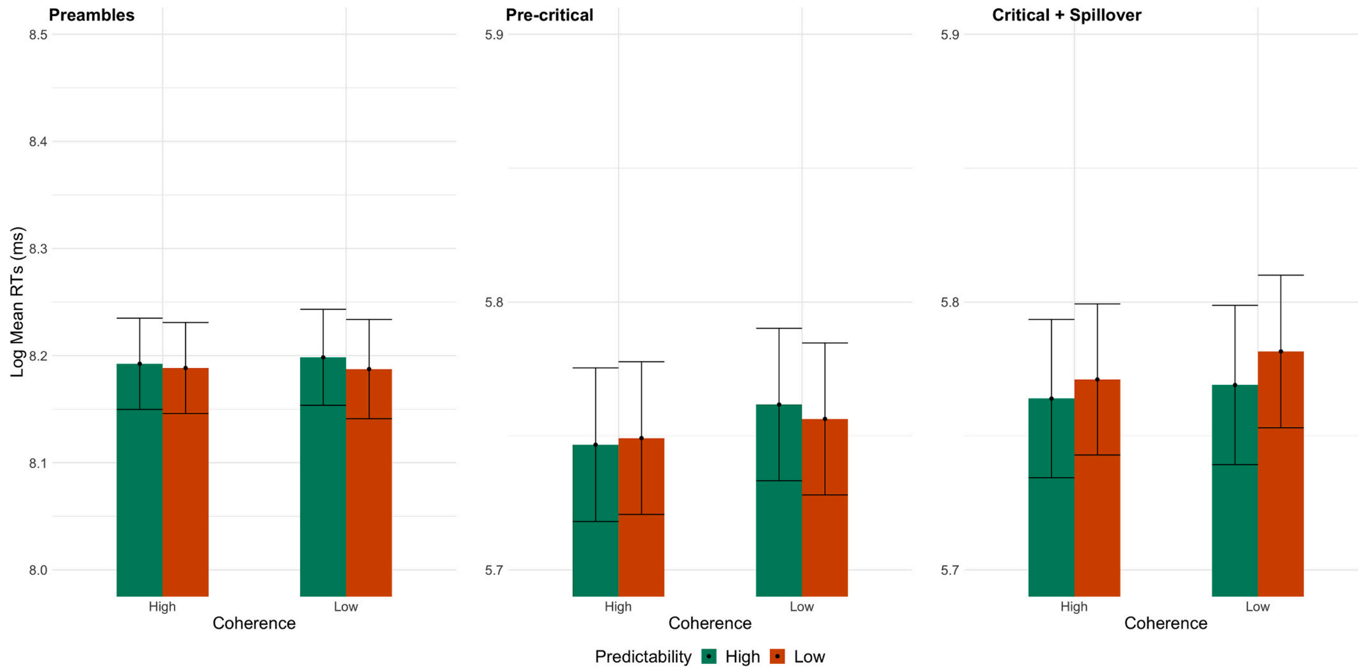

Fig. 2. Experiment 1 pre-registered analysis results; error bars indicate the 95% confidence intervals.

−1.62, $p = 0.11$). Further, we found no interaction effect (β = 0.001, SE = 0.002, $t = 0.54$, $p = 0.59$). We did find marginal effects of critical word length (β = 0.007, SE = 0.004, $t = 1.94$, $p = 0.05$) and word frequency (β = −0.006, SE = 0.003, $t = -1.80$, $p = 0.07$).

2.2.1.4. Critical word. At the critical word, there was no processing advantage for words presented with high coherent preambles, compared to less coherent preambles (β = −0.003, SE = 0.002, $t = -1.47$, $p = 0.14$). There was no effect of predictability (β = −0.005, SE = 0.003, $t = -1.61$, $p = 0.11$), nor was there an interaction (β = 0.002, SE = 0.003, $t = 0.65$, $p = 0.52$). We did find marginal effects of word length (β = 0.007, SE = 0.004, $t = 1.81$, $p = 0.07$), and frequency (β = −0.007, SE = 0.003, $t = -1.96$, $p = 0.05$).

In sum, for our pre-registered analyses for Experiment 1, we found that coherence facilitated processing at the pre-critical ROI, such that faster reading times were observed for high coherent trials compared to less coherent trials. However, this effect did not extend into our main ROI, the critical word and spillover region.

2.2.2. Exploratory analyses

After completing our pre-registered analyses, we made a number of changes to our analysis pipeline to (a) improve our data exclusion criteria, (b) include more rigorous data pre-processing steps (e.g., residual reading times) and (c) treat coherence as a continuous variable. The modified analysis protocol used here was then pre-registered for Experiments 2 and 3.

Upon inspection of the data, we found that a small number of participants had very short, long or variable reading times for the preambles, suggesting that they were not paying sufficient attention to them. Hence, we added an additional exclusion criterion associated with preamble RTs. If mean RTs for the preambles was either <2 s or >15 s ($n = 11$), or if the standard deviation of a participant's preamble RTs was >15 s ($n = 6$), we excluded that participant from the analyses. Finally, we removed all trials whose comprehension questions had a low mean accuracy across all participants (<70%), indicating that there was a general difficulty in understanding the sentence ($n = 7$).

To account for effects of word length and individual differences in reading speeds, we fit an initial linear model on the log RTs from all words and ROIs, with a fixed effect of word length and random-intercepts by participant. We then used the residuals from this linear model as the dependent variable in our main analysis models (Enochson & Culbertson, 2015; Ferreira & Clifton, 1986). The use of residual reading times to account for differences in word length and differences in individual participant's reading speeds has become widely accepted for self-paced reading paradigms (Lorch & Myers, 1990; Trueswell, Tanenhaus, & Garnsey, 1994). For ROIs that consisted of more than one word, we took the average residual log RT per trial.

Further, we transformed our variable of interest, coherence, from a two-level factor to a continuous predictor. We took the coherence ratings acquired during the prior norming experiment and scaled these. This gave a finer-grained representation of coherence, rather than binning it into two arbitrary groups. It also allowed us to fully disaggregate the effect of predictability from that of coherence (as HP trials had slightly higher coherence values than LP trials).

Finally, we conducted additional analyses including possible confounds as fixed effects. First, we included experiment half (first vs. second) and its interactions with predictability and coherence, to investigate whether participants' behaviour changed during the course of the experiment. Second, we included a binary fixed effect coding of whether or not the critical word was also present in the preamble for a trial. Word repetition occurred for only a handful of the 800 variations of our stimuli; however previous research has demonstrated robust repetition priming effects on RTs (Forster & Davis, 1984). The inclusion of these additional covariates had no effect on the main results of any model; thus we report the results of the models without the additional factors below.

2.2.2.1. Preambles. Estimated means for the new analyses are shown in Fig. 3. For the preambles ROI, as expected, we found no advantage in reading times for highly coherent preambles (β = 0.002, SE = 0.01, $t = 0.22$, $p = 0.82$) or for trials with highly predictable critical words (β = 0.003, SE = 0.005, $t = 0.60$, $p = 0.55$), nor an interaction between the two (β = 0.006, SE = 0.009, $t = 0.64$, $p = 0.52$).

2.2.2.2. Pre-critical word. For the pre-critical ROI, participants read more coherent trials significantly faster than less coherent trials (β = −0.009, SE = 0.003, $t = -2.98$, $p = 0.003$). As expected, there was no advantage for trials with highly predictable critical words (β = 0.001,

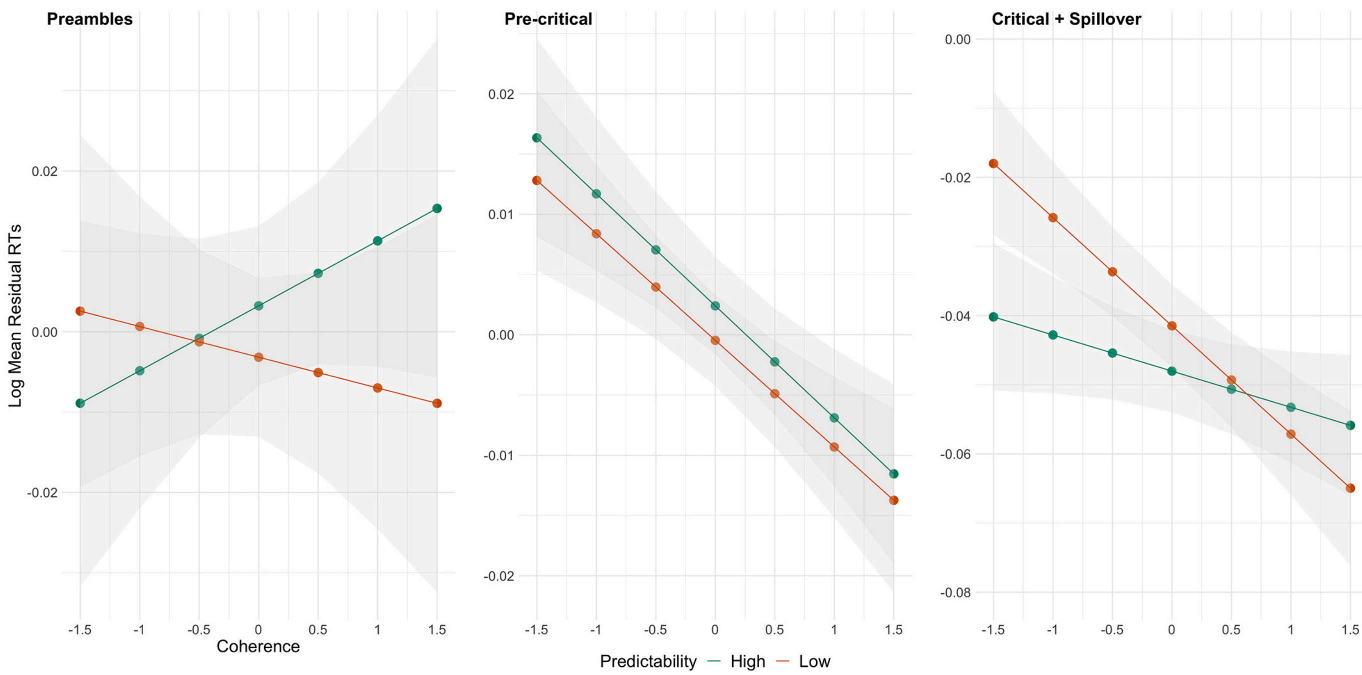

Fig. 3. Experiment 1 exploratory analysis results; shaded areas depict the standard error.

SE = 0.001, t = 0.86, p = 0.39) and no interaction between coherence and predictability (β = −0.0002, SE = 0.003, t = −0.07, p = 0.94).

2.2.2.3. Critical word and spillover. For this ROI, we found that participants' reading times were affected by coherence, such that more coherent preambles facilitated language processing (β = −0.01, SE = 0.004, t = −2.47, p = 0.01). However, we found no difference in reading times between high and less predictable critical words (β = −0.003, SE = 0.002, t = −1.41, p = 0.16), nor was there an interaction between the two (β = 0.005, SE = 0.004, t = 1.29, p = 0.20). Further, there was no effect of critical word frequency on reading times (β = 0.0001, SE = 0.003, t = 0.03, p = 0.97).

2.2.2.4. Critical word. When considering the critical word only, we found that critical words preceded by more coherent preamble contexts had a processing advantage (β = −0.009, SE = 0.004, t = −2.22, p = 0.03). However, there was no facilitation of highly predictable critical words (β = −0.004, SE = 0.003, t = −1.55, p = 0.12). Reading times were not affected by the Zipfian frequency of the critical word (β = −0.003, SE = 0.003, t = −0.86, p = 0.39), nor by an interaction between coherence and predictability (β = 0.006, SE = 0.005, t = 1.24, p = 0.22).

2.3. Discussion

Experiment 1 found that people were faster to read target sentences when they occurred after a highly coherent context. This effect emerged at the pre-critical ROI, which is where the lack of coherence between the preambles and target sentence would begin to become apparent, and (in our improved analysis pipeline) extended into the critical word and spillover ROI. Thus, it appears that contextual support from the global discourse impacted reading times to a greater degree than contextual support from the local sentence in this experiment.

The absence of a significant predictability effect at the critical word and spillover ROI is also notable. We identify three possible reasons as to why this effect was weaker than expected. First, with the addition of a global context, it is possible that information from this takes precedence over information from the local (sentence) context, as suggested by some previous studies (Hess et al., 1995; Ledoux et al., 2006). Thus, local predictability may have little influence in this situation. An alternative explanation is that the wider experimental context discourages predictive processing due to the frequent occurrence of unexpected stimuli: 75% of trials in the experiment contained at least one kind of expectation violation (either a lack of coherence between preamble and target sentence, the presence of a less predicted critical word, or both). This conclusion would be consistent with Brothers et al.'s (2017) proposal that participants reduce their use of predictive processing when in an experimental environment in which expectations are frequently disconfirmed. Finally, because we used low predictable target words that were nevertheless semantically plausible for their preceding context, the predictability effect may be small and thus we may have been unable to detect it.

We found no hint of an interaction effect between coherence and predictability. This could indicate that local expectations are not influenced by the presence of discourse-level incoherence, which would suggest a more automatic mechanism for predictive processing. However, it is difficult to conclude this without observing a reliable main effect of predictability. Thus, we designed Experiments 2 and 3 to test whether an interaction effect would be present in conditions that promoted larger effects of predictability. In Experiment 2, we altered the ratio of high and less coherent contexts across the experiment to a 75:25 split, in order to reduce the overall frequency with which expectations were violated and address this potential explanation for the absence of the predictability effect. In Experiment 3, we replaced the low predictability critical words (which, though not predicted, were semantically plausible) with anomalous words that violated the semantic expectations provided by the target sentence. This was intended to induce the strongest possible predictability effect, thereby addressing another possible explanation for its absence.

3. Experiment 2

The purpose of Experiment 2 was to address the possibility that the lack of predictability effects in Experiment 1 was due to the even ratio of more and less coherent contexts. It is possible the high prevalence of incoherent narratives led participants to disengage from predictive processing during the experiment as a whole (Brothers et al., 2017).

Incoherent passages are relatively infrequent in natural language contexts, with interlocutors typically maximising the amount of coherence in their message (Black, 1988; Grice, 1975, 1989). Thus, to determine whether the same results would be found under conditions where topic shifts occurred less frequently, we reduced the frequency of less coherent passages from 50% to 25% of trials.

3.1. Methods

3.1.1. Participants

Participants were recruited through the Testable Minds online participant pool. We required that the participants were based in the UK, had an approval rating >90%, were native speakers of English and had a maximum age of 40. They were reimbursed $5.60 for their time. In total, we collected data from 134 participants (mean age: 28, range: 18–40, SD = 6.09). 26 participants were excluded for failing to meet our pre-registered performance criteria (see below), leaving 108 participants in the final analysis.

3.1.2. Stimuli

The stimuli used in Experiment 2 were identical to Experiment 1; however, we altered the ratio of trials with more and less coherent preambles in order to reduce the amount of topic shifts that occurred across the experiment as a whole. This ratio was changed to 75:25 for more and less coherent contexts. To do so, we took the experimental lists from Experiment 1 and replaced half of the LC trials with their equivalent HC preambles. We ensured that each block (every 40 trials) had the same ratio of HC and LC trials (15 for each HC and 5 for each LC, respectively). We also ensured that each trial appeared in all four conditions across the experimental lists.

3.1.3. Procedure

The procedural steps for Experiment 2 were identical to Experiment 1, except for one change to the presentation of the preambles. We ensured that preambles appeared on-screen for a minimum of two seconds before advancing to the target sentence, irrespective of how quickly participants responded (unlike in Experiment 1 where participants could advance in <2 s). This was to encourage participants to process the preambles more fully. Because of this, we did not log transform RTs for this ROI (unlike the exploratory analyses in Experiment 1).

3.1.4. Analysis

As with Experiment 1, we pre-registered our analyses prior to data collection (https://osf.io/zu76h). Our pre-registered analyses for Experiment 2 used the analysis approach from our exploratory analyses of Experiment 1 (except for log-transforming RTs in the preambles ROI). In terms of exclusion criteria, participants were excluded if their comprehension accuracy was lower than 80% ($n = 21$), and if the SDs of their preamble RTs was >15 s ($n = 5$). Trials with incorrect answers to the comprehension questions were also removed ($n = 440$).

3.2. Results

3.2.1. Pre-registered analyses

3.2.1.1. Preambles. For the preambles (see Fig. 4), we found no differences in reading times for more and less coherent conditions ($\beta = -29.65$, SE = 54.70, $t = -0.54$, $p = 0.59$), no differences in reading times between high and less predictable conditions ($\beta = -36.02$, SE = 37.35, $t = -0.96$, $p = 0.34$), nor any interaction ($\beta = -30.40$, SE = 51.89, $t = -0.59$, $p = 0.56$).

3.2.1.2. Pre-critical word. At the pre-critical word region, we found a marginal effect of coherence, such that there was a suggestion that participants may have read more coherent conditions faster than less coherent conditions ($\beta = -0.005$, SE = 0.003, $t = -1.74$, $p = 0.08$). As expected, we found no impact of predictability on reading times ($\beta = -0.001$, SE = 0.002, $t = -0.73$, $p = 0.46$), nor an interaction between the two ($\beta = -0.001$, SE = 0.003, $t = -0.42$, $p = 0.68$).

3.2.1.3. Critical word and spillover. At our main ROI, we found no influence of coherence ($\beta = -0.01$, SE = 0.01, $t = -1.46$, $p = 0.15$), predictability ($\beta = -0.001$, SE = 0.004, $t = 0.25$, $p = 0.80$), or their interaction ($\beta = -0.005$, SE = 0.009, $t = -0.56$, $p = 0.58$). Further, there was no effect of critical word frequency ($\beta = 0.003$, SE = 0.004, $t = 0.79$, $p = 0.43$).

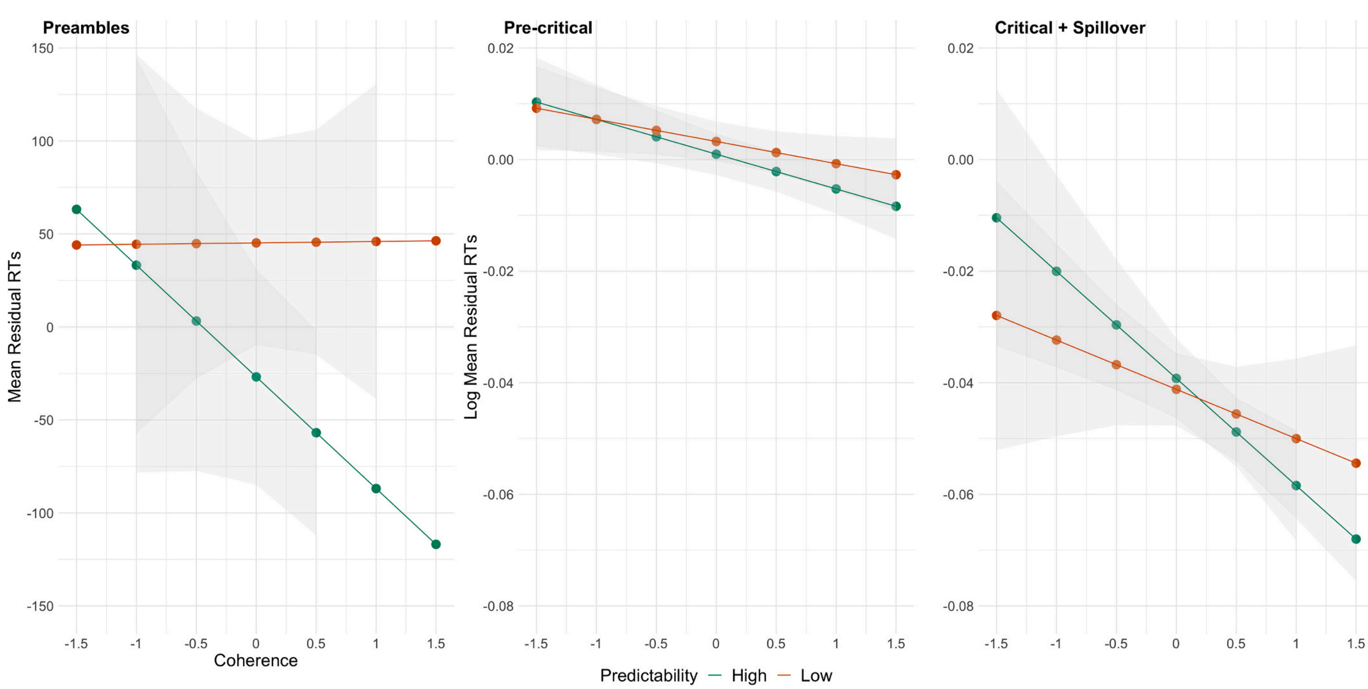

Fig. 4. Experiment 2 pre-registered results; shaded areas denotes standard error.

3.2.1.4. Critical word. At the critical word alone, we found no differences in reading times between more and less coherent conditions (β = −0.01, SE = 0.009, $t = -1.38$, $p = 0.17$), high and less predictable conditions (β = −0.001, SE = 0.003, $t = -0.42$, $p = 0.67$), nor their interaction (β = −0.002, SE = 0.008, $t = -0.24$, $p = 0.81$). Further, there were no differences in reading times for high and less frequent critical words (β = −0.001, SE = 0.004, $t = -0.32$, $p = 0.75$).

3.2.2. Combined experiment 1 and 2

In order to directly compare the effects from Experiments 1 and 2, we combined the datasets and ran additional linear mixed-effects models, including fixed effects of experiment and its interactions with coherence and predictability.

3.2.2.1. Preambles. For the preambles region, there was no effect of experiment or an interaction between experiment and the other predictors (all $p > 0.05$).

3.2.2.2. Pre-critical word. For this ROI, we found no effect of experiment and there was no interaction with the other predictors (all $p > 0.05$).

3.2.2.3. Critical word and spillover. At our main ROI, we found a marginal effect of coherence (β = −0.009, SE = 0.005, $t = -1.78$, $p = 0.08$). There was no effect of predictability (β = −0.002, SE = 0.002, $t = -1.16$, $p = 0.25$), no interaction (β = 0.0006, SE = 0.003, $t = 0.21$, $p = 0.84$), and no effect of critical word frequency (β = −0.001, SE = 0.002, $t = -0.46$, $p = 0.65$). Further, there was no effect of experiment or an interaction between experiment and the other predictors (all $p > 0.05$).

3.2.2.4. Critical word. At this region, we found marginal effects of coherence (β = −0.009, SE = 0.005, $t = -1.86$, $p = 0.07$), predictability (β = −0.003, SE = 0.002, $t = -1.72$, $p = 0.09$) and log frequency (β = −0.004, SE = 0.002, $t = -1.79$, $p = 0.07$). We found no difference in reading times for the interaction of coherence and predictability, no effect of experiment nor any interactions of experiment with the other predictors (all $p > 0.05$).

3.3. Discussion

Experiment 2 attempted to test whether the predictability effect (and as an extension, its interaction with coherence) would occur in a more naturalistic experimental environment, where incoherent passages occurred less frequently. However, we found that even after adjustments to the experimental environment to promote predictive processing, still no predictability effect was found. Further, there was no interaction effect with coherence, replicating our findings from Experiment 1. Effects of coherence at the target sentence (emerging at the pre-critical ROI) did not reach statistical significance in this experiment, though they were of a similar magnitude to those observed in Experiment 1 and a combined analysis indicated that the size of the coherence effects did not differ significantly between experiments. We suggest that the lack of a significant coherence effect here may be due to reduced power as less coherent passages were sampled less frequently. For Experiment 3, we returned to an even ratio of high and low coherence trials and investigated the effect of including semantically anomalous (as well as unpredictable) critical words.

4. Experiment 3

Experiment 3 was conducted in order to address the absence of a predictability effect in Experiments 1 and 2. It is possible that an effect of predictability did not appear due to the nature of our critical words. The less predictable conditions in Experiments 1 and 2 were created using a critical word with a lower predictability than the target critical word (using the data from Peelle et al., 2020), but in a similar semantic domain to the high predictable critical word. They therefore represented a violation of expectations at the lexical level but not the semantic level. It is possible that this manipulation was not strong enough to generate a reliable predictability effect, due to the high degree of semantic feature overlap between the less predictable (but semantically plausible) continuation and the expected continuation (Federmeier & Kutas, 1999). As such, in Experiment 3, we replaced less predictable critical words with semantically anomalous words, which disconfirmed expectations at both the lexical and semantic levels. Thus, we expected this manipulation to produce more reliable effects of predictability, increasing our sensitivity to detect an interaction of this effect with coherence.

4.1. Methods

4.1.1. Participants

Participants were recruited through the online platform, Prolific. We filtered the participant pool using the following eligibility criteria: 1) native speakers of English, 2) currently located within the UK, 3) a maximum age of 40, and 4) a minimum approval rating of 90%. Participants were reimbursed with £5.63 for their time. In total, we collected data from 126 participants (mean age: 28, range: 18–40, SD: 5.75). Our final sample size consisted of 97 participants ($n = 4$ removed for a SD of their preambles RTs larger than 15 s; $n = 25$ removed for having a comprehension accuracy lower than 80%).

4.1.2. Stimuli

We created two new conditions for each trial that included an anomalous critical word. We replaced the less predictable stimuli with these anomalous stimuli. An example trial with all four conditions can be found in Table 3.

We ensured that the anomalous critical words had the same part-of-speech as the high and low predictable critical words, ensuring that the violation would be purely lexico-semantic. Moreover, we matched the anomalous critical words on length ($t = 0$, $p = 1$), Zipfian frequency ($t = 0.29$, $p = 0.77$) and bigram frequency ($t = -1.88$, $p = 0.06$) with the highly predictable critical words (see Table 4). All of the anomalous critical words had a cloze value of 0 within the local context of the target sentence ($t = -165.04$, $p < 0.001$). We used the same high and low coherent preambles as Experiments 1 and 2, therefore the matching on preamble length was the same ($t = 1.11$, $p = 0.27$).

We did not obtain new coherence ratings for passages containing anomalous words as we assumed that the presence of a strong semantic violation in the target sentence would lead raters to give a low coherence rating to all of these stimuli, irrespective of how coherent the discourse was prior to the violation. This would be problematic as we were interested in how the coherence of the discourse leading up to the critical word influences its processing. Instead, we took the coherence ratings from HC/HP and LC/HP versions of the stimuli and used these as the coherence values in the HC/Anom and LC/Anom conditions. This

Table 3
Example sentences across the four conditions in Experiment 3.

Condition	Context sentences	Target sentence
High Coherence, High Predictability	Crime was a problem. Residents were starting to feel unsafe.	The police arrested the local gang for selling **drugs** on the corner.
High Coherence, Anomalous		The police arrested the local gang for selling **words** on the corner.
Low Coherence, High Predictability	The city was a safe place. It had a low crime rate and high employment.	The police arrested the local gang for selling **drugs** on the corner.
Low Coherence, Anomalous		The police arrested the local gang for selling **words** on the corner.

Table 4
Average psycholinguistic properties of conditions in Experiment 3.

Property	HC/HP	HC/Anom	LC/HP	LC/Anom
Coherence	3.83 (0.46)	3.83 (0.46)	2.46 (0.52)	2.46 (0.52)
Preamble Length	13.54 (2.05)	13.54 (2.05)	13.78 (1.92)	13.78 (1.92)
Critical Word Length	4.35 (1.16)	4.35 (0.90)	4.35 (1.16)	4.35 (0.90)
Critical Word Zipf Frequency	4.53 (0.80)	4.56 (0.68)	4.53 (0.80)	4.56 (0.68)
Critical Word Bigram Frequency	1794.57 (1536.06)	1504.95 (1087.85)	1794.57 (1536.06)	1504.95 (1087.85)
Cloze Value	86.23 (6.59)	0.00 (0.00)	86.23 (6.59)	0.00 (0.00)

Note. Standard deviations are shown in parentheses. HC = high coherence; HP = high predictability; LC = low coherence; Anom = anomalous predictability.

means that HC/HP and HC/Anom have the same coherence values in our analyses, reflecting the fact that these conditions are identical up to the critical word (and the same for LC/HP and LC/Anom). As in Experiments 1 and 2, coherence differed significantly between HC and LC stimuli ($F = 1238.19$, $p = 2.19 \times 10^{-151}$).

4.1.3. *Procedure*

The procedural steps for Experiment 3 were identical to Experiment 2, except that we returned to presenting each participant with an equal number of HC and LC trials.

4.1.4. *Analysis*

Similar to Experiments 1 and 2, we pre-registered our analyses prior to data collection (https://osf.io/cqr7a). All of our analytical steps and exclusion criteria are identical to Experiment 2.

4.2. *Results*

4.2.1. *Pre-registered analyses*

4.2.1.1. Preambles. Model estimates of reading times are presented in Fig. 5. For the preambles, as expected, we found no difference in reading times for the effect of coherence (β = −36.05, SE = 72.97, $t = -0.49$, $p = 0.62$), no difference in reading times for high and anomalous predictable conditions (β = 46.10, SE = 36.74, $t = 1.26$, $p = 0.21$), nor any interaction (β = −82.74, SE = 57.97, $t = -1.43$, $p = 0.15$).

4.2.1.2. Pre-critical word. At the pre-critical word region, we found no effect of coherence (β = −0.00003, SE = 0.003, $t = -0.01$, $p = 0.99$), no differences between high and anomalous predictable conditions (β = −0.001, SE = 0.002, $t = -0.82$, $p = 0.41$), nor any evidence of an interaction (β = 0.004, SE = 0.003, $t = 1.33$, $p = 0.18$).

4.2.1.3. Critical word and spillover. At our main ROI, we found a significant effect of predictability, such that participants read the critical word and spillover region faster for highly predictable critical words than for anomalous critical words (β = 0.005, SE = 0.002, $t = 2.25$, $p = 0.02$). We also found a significant interaction between coherence and predictability (β = 0.009, SE = 0.004, $t = 2.13$, $p = 0.04$), such that the facilitation for highly predictable words was greater on more coherent trials, than for anomalous words. In post-hoc analyses, we ran separate linear mixed-effects models for the high coherence and low coherence trials. A predictability effect was present for the more coherent trials (β = 0.007, SE = 0.003, $t = 2.18$, $p = 0.03$), but not for the less coherent trials (β = 0.004, SE = 0.003, $t = 1.26$, $p = 0.20$). Finally, we found no main effect of coherence (β = 0.003, SE = 0.004, $t = 0.60$, $p = 0.55$), nor an effect of frequency (β = −0.004, SE = 0.003, $t = -1.39$, $p = 0.16$).

4.2.1.4. Critical word. Analyses of the critical word alone produced similar results to those that included the spillover region. There were significant reading time differences between the highly predictable and anomalous conditions (β = 0.005, SE = 0.002, $t = 2.10$, $p = 0.04$). Further, the interaction of coherence and predictability was marginally significant (β = 0.008, SE = 0.004, $t = 1.91$, $p = 0.06$), as was the impact of critical word frequency (β = −0.005, SE = 0.003, $t = -1.77$, $p = 0.08$). No differences in reading times were found for the high and low coherent conditions (β = 0.003, SE = 0.004, $t = 0.59$, $p = 0.56$).

4.2.2. *Combined experiment 1 and 3*

To compare the results from Experiments 1 and 3, we also ran further analyses on a combined dataset from these experiments. Analyses

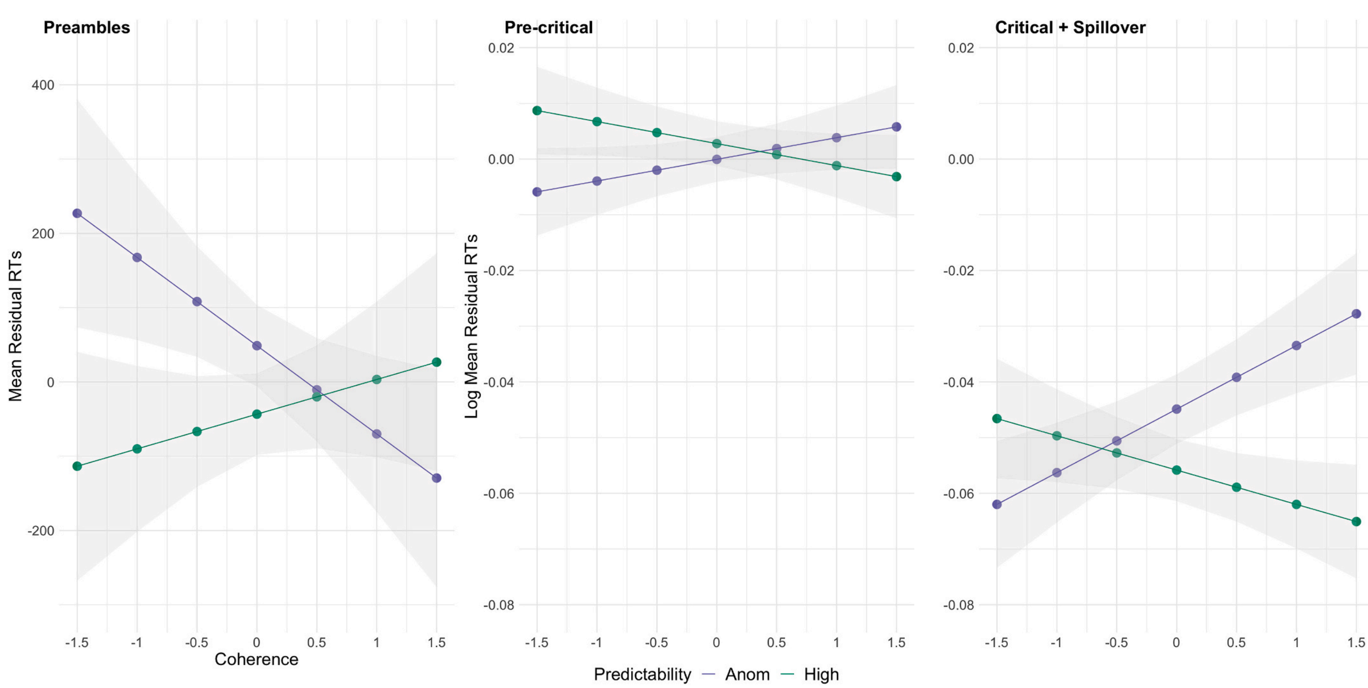

Fig. 5. Experiment 3 pre-registered results; shaded areas indicate the standard error.

included fixed effects of experiment and its interactions with coherence and predictability.

4.2.2.1. Preambles. At this ROI, we found a significant interaction of predictability and experiment (β = 0.006, SE = 0.003, t = 2.06, p = 0.04) (see Fig. A1 in Supplementary Materials). This is likely to be a false positive as the low and high predictability trials had the same preambles across both experiments. No other results reached significance, including the test of experiment and its interactions (all $p > 0.05$).

4.2.2.2. Pre-critical. For the pre-critical region, less coherent trials were read more slowly (β = −0.004, SE = 0.002, t = −2.07, p = 0.04), and this effect interacted with experiment (β = −0.004, SE = 0.002, t = −2.15, p = 0.03), such that the coherence effect was larger in Experiment 1 (as shown in Fig. 6). This is likely due to the fact that a coherence effect was observable in Experiment 1, and not in Experiment 3. There were no other main effects or interactions (all $p > 0.05$).

4.2.2.3. Critical word and spillover. For our main ROI, we found that anomalous/less predictable words were read slower (β = −0.005, SE = 0.002, t = −2.76, p = 0.006) than highly predictable words and critical words with higher frequency were read faster (β = −0.004, SE = 0.002, t = −2.02, p = 0.04). Moreover, there was an interaction between coherence and experiment (β = −0.007, SE = 0.003, t = −2.25, p = 0.02). Coherence had a larger overall impact on reading times in Experiment 1 than in Experiment 3, as an overall coherence effect was not observed in Experiment 3. Importantly, a three-way interaction emerged between coherence, predictability and experiment (β = 0.007, SE = 0.003, t = 2.53, p = 0.01). This indicates that the pattern observed in Experiment 3, whereby the predictability of the critical word influenced processing only in more coherent passages, differed significantly from that of Experiment 1 (see Fig. 6 for illustration). There were no other main effects or interactions (all $p > 0.05$).

4.2.2.4. Critical word. At the critical word alone, we found that anomalous/less predictable words were read more slowly than highly predictable words (β = −0.005, SE = 0.002, t = −2.93, p = 0.004), as well as that critical words with higher log frequency were read faster than those with lower log frequency (β = −0.007, SE = 0.002, t = −3.16, p = 0.001). Further, a three-way interaction between coherence, predictability and experiment emerged (β = 0.006, SE = 0.003, t = 2.13, p = 0.03), and a marginal interaction between coherence and experiment was present (β = −0.005, SE = 0.003, t = −1.81, p = 0.07). We found no other main effects or interactions (all $p > 0.05$).

4.3. Discussion

In Experiment 3, we replaced the less predictable critical words with semantically anomalous words that violated readers' expectations at the semantic as well as the lexical level. We expected to observe larger effects associated with the critical word, and we did: there was a main effect of predictability, whereby people were slower to read anomalous critical words, compared with highly predictable critical words. However, the size of this effect varied according to discourse coherence, with anomalous critical words slowing processing only when preceded by more coherent contexts. In other words, the more coherent global context facilitated the processing of highly predictable critical words, but impaired the processing of anomalous critical words. We interpret these results as evidence of variation in the engagement of predictive processes. Reading of highly predictable critical words was facilitated when these were preceded by a highly coherent context, suggesting that these conditions encouraged reliance on expectations. At the same time, reading of anomalous words was slower in highly coherent contexts, consistent with the need to overcome the effect of strong but incorrect expectations. Thus, these results suggest that the degree to which predictive processing is engaged depends on the surrounding context. It is possible that comprehenders downregulate predictive processing when the global context is not informative, thus reducing the effects at the level of the local sentence.

Another interpretation of the interaction effect is that it relates to variations in how long readers spend verifying that a word is anomalous. Specifically, it is possible that more time is taken to confirm whether a word is semantically anomalous when it is embedded in a more coherent discourse, as the passage made sense until that point. In other words, the anomaly may be more surprising on high coherence trials. It is difficult to disentangle this possibility from a predictive processing account, as

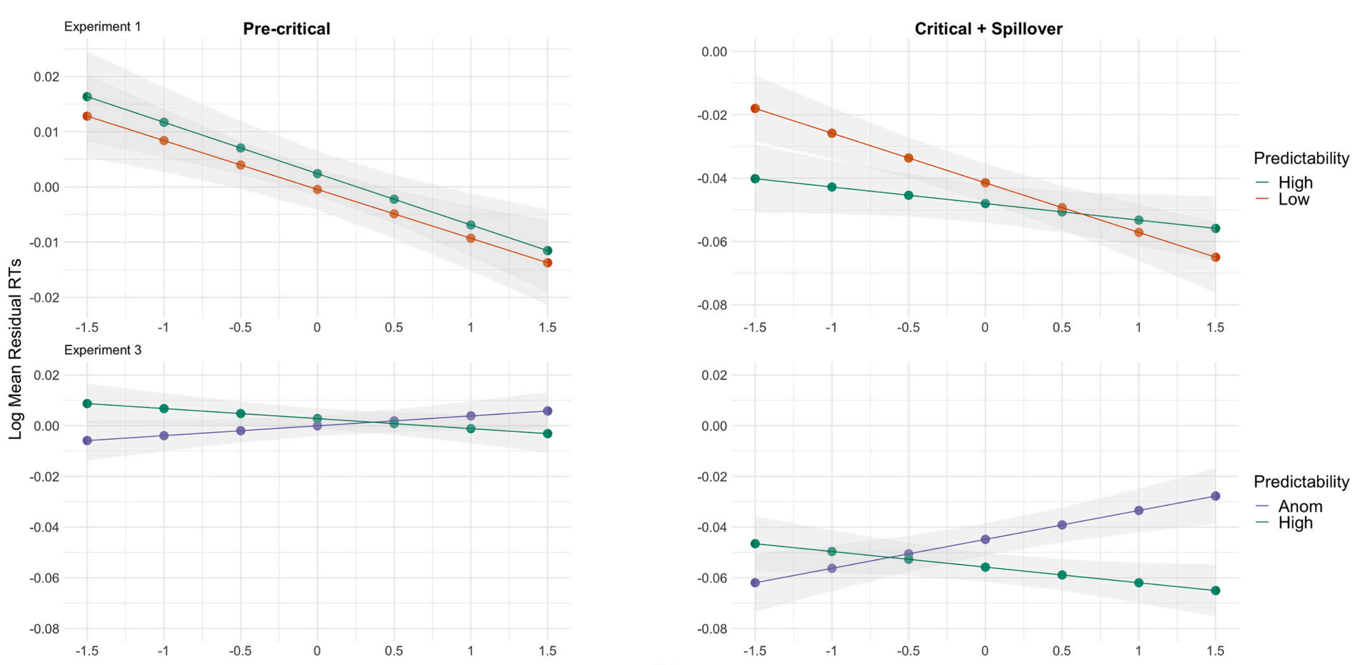

Fig. 6. Results for pre-critical word and critical word and spillover regions for Experiment 1 (top) and Experiment 3 (bottom); shaded areas denote standard error.

both are rooted in readers' expectations about the critical word. In addition, an explanation based solely on anomaly processing does not provide an explanation for facilitation of predictable words, which we also observed.

5. General discussion

The current study addressed how discourse coherence influences predictive processing during language comprehension. We were particularly interested in how the validity of the global discourse context affected comprehenders' use of local sentence constraints. To this end, we conducted three online, self-paced reading experiments using three-sentence discourses. Our stimuli were manipulated on both the predictability of the critical word given its local sentence, and the coherence of the target sentence, given the preceding preamble. Results from Experiment 1 suggested tht global discourse had an influence on processing, with a coherence effect emerging at the pre-critical ROI, and extending into the critical word and spillover ROI. This suggests that information from the global discourse context is used to facilitate comprehension. However in Experiment 1, while discourse coherence benefited processing, no local predictability effects were observed. In our second experiment, we used a lower proportion of less coherent trials, but found similar results to Experiment 1. Experiment 3 used anomalous critical words that more strongly violated comprehenders' expectations. Here, an effect of predictability did emerge, with facilitation for highly predictable words compared to anomalous words. Importantly, however, this effect was only observed when the target sentence was embedded in a highly coherent passage, with no predictability advantage present for less coherent discourse. These results indicate, first, that relatively subtle manipulations of discourse coherence influence the speed of comprehension and second, that a reduction in the coherence of a discourse appears to reduce the effects of local sentence constraints. This latter effect is in line with the proposal that comprehenders use a range of environmental cues to regulate their use of predictive processing. We will discuss each of these effects in turn.

5.1. Influence of coherence

In our first two experiments, the impact of coherence on how comprehenders process upcoming material was clear. Trials preceded by more coherent preamble contexts were read faster than those preceded by less coherent contexts. In Experiment 1, this coherence effect emerged at the pre-critical region and extended into the critical word and spillover region. The emergence of the coherence effect at the pre-critical region indicates the rapid influence of cues from global context on reading times. The pre-critical ROI is the first region where the lack of coherence between the preamble context and the target sentence could be detected. While the coherence effect in Experiment 2 did not reach statistical significance, possibly due to the lower sampling of less coherent contexts, the numerical effect was of a similar magnitude as in Experiment 1. These results hint at the privileged position information from the global context plays in reading comprehension. Our results are consistent with previous findings that the presence of a highly coherent discourse context facilitates comprehension, while deviation from the existing discourse model results in a slowdown (Albrecht & O'Brien, 1993; Hess et al., 1995; Myers & O'Brien, 1998; O'Brien & Albrecht, 1992; Schwanenflugel & White, 1991; Stewart, Kidd, & Haigh, 2009).

We theorise that comprehenders initially build a discourse model of the events portrayed in the preamble contexts, then when faced with a target sentence that is not coherent with this established mental model, processing difficulties emerge. The mismatch between the established mental model and the incoming incoherent information can be established early into the target sentence and, indeed, we found that the coherence effect emerges at the pre-critical region. These processing difficulties appear to be sustained in time, such that they impact the processing of later words within the target sentence. This suggests that the comprehender experiences continued difficulties in integrating the incoming information into their established mental model of the discourse. The prolonged slowing in reading times could be indicative of extended and perhaps repeated attempts to re-update the mental model as new information becomes available (Albrecht & O'Brien, 1993; Kuperberg & Jaeger, 2016; Van Petten & Luka, 2012). One interesting line of future research would be to have a finer-grained measure of when, and for how long, the coherence effect impacts processing. One way in which this could be achieved would be to track the emergence of coherence effects at the single word level, using either RTs or ERP measures, to determine when disruptions to coherence become apparent to comprehenders. Predictions from NLP models (Frank, Otten, Galli, & Vigliocco, 2013; Szewczyk & Federmeier, 2022) could inform this effort by providing information on how breaks in coherence influence the surprisal associated with words in the target sentence.

Surprisingly, in our third experiment, we did not find an effect of coherence. This was unexpected as the stimuli were the same as those presented in Experiment 1 up until the anomalous critical word. As such, we would have expected a similar coherence effect to emerge at the pre-critical ROI across all experiments. The fact that this is absent in Experiment 3 suggests that a difference in the experimental environments (between Experiments 1 and 3) may have induced changes in processing strategy. In Experiment 1, unpredictable critical words were from the same semantic domain as predictable critical words, so they could be accommodated by updating the discourse model. In contrast, Experiment 3 used semantically anomalous critical words that were entirely incompatible with any meaningful interpretation of the discourse. It is possible that participants became aware that on many trials they would encounter an anomalous critical word, and that their discourse model would fail. This could have led participants to delay integration of the target sentence with the preambles until later in the sentence to avoid wasted effort. Our current findings are not able to concretely conclude this, so further study on this is warranted.

5.2. The role of predictability

The first two experiments failed to observe a main effect of predictability. In these experiments, we created our less predictable critical word condition by swapping the highly predictable critical word with a lower cloze continuation of a matching syntactic role that still remained in the same semantic domain. Thus, while the specific lexical form of our less predictable critical words was not expected, they were still compatible with expectations at the syntactic and semantic levels and may have been facilitated to some extent (Amsel, DeLong, & Kutas, 2015; Federmeier & Kutas, 1999; Metusalem et al., 2012; Wlotko & Federmeier, 2012). To address this, we used semantically anomalous critical words in Experiment 3 (replacing the less predictable condition). Here, we found that when coherence was high, comprehenders were faster to read the highly predictable continuations than the anomalous targets. This was as expected: all of the target sentences allowed comprehenders to generate a strong expectation about the critical word, and processing of this word was then facilitated when it appeared. However, if a semantically anomalous word was observed instead, processing became more difficult as the prior expectation was unhelpful (and disconfirmed). Critically, however, this predictability effect did not occur when the target sentence appeared in a less coherent discourse. On these trials, the contents of the target sentence were the same, and therefore in principle, would still lead to an expectation that favours the highly predictable continuation. However, no such facilitation was observed. In addition, reading of anomalous words was faster when they appeared in less coherent passages, suggesting that no prediction was disconfirmed when coherence was low. Thus, it appears that when the preamble contexts were less coherent, local sentential constraints no longer affected processing to the same degree as for more coherent contexts.

These results add to the evidence that expectations about upcoming language material are recruited flexibly and are influenced by a range of contextual cues beyond the immediate sentence (Brothers, Wlotko, Warnke, & Kuperberg, 2020; Huettig, 2015; Kuperberg & Jaeger, 2016). For example, previous studies have shown that the degree of facilitation for predictable words depends on the comprehender's knowledge of the speaker and the frequency with which predictions are violated in the experiment (Brothers et al., 2017, 2019). Thus, when people have reason to believe that their predictions will not be valid, they appear to pre-activate predictable words less. The present results indicate that the coherence of the current discourse is another cue that comprehenders use to regulate their use of predictive processing. On less coherent trials, we propose that as participants processed the target sentence, they became aware that the mental model established during the preamble was no longer valid. The introduction of new information that was incompatible with the existing mental model may have acted as a signal that expectations were likely to be unreliable. Thus, the language system disengaged attempts to use expectations to facilitate processing. This would explain the absence of an advantage for highly predictable continuations over semantically anomalous words.

Importantly, our effects occurred even though the target sentence itself was still highly constraining towards a particular completion. The uncertainty regarding the state of the global discourse seemed to prevent these constraints from being used to aid processing. We also highlight that coherence itself was not a reliable cue that the target sentence would contain a violation, since violations were equally common in high and low coherence trials. As such, we do not think these effects reflect a specific strategy used in our experiment. Instead, the downregulation of predictive processing appears to be a natural response to detecting that the discourse is in a state of uncertainty.

Our results are in opposition to those found by Camblin et al. (2007), whereby facilitation from local within-sentence associations were largest when preceded by extremely incoherent contexts. In that study, the incoherent contexts consisted of sentences scrambled between different trials, so they were completely uninformative and no discourse model could be constructed. We suggest that the difference between the current study and Camblin et al. (2007) is that when the global context is completely uninformative, comprehenders disregard it and rely entirely on information at the local level. In the current study, however, the more subtle manipulation of coherence may have led comprehenders to attempt to form a coherent representation of the passages. In this situation, we suggest that disruption to coherence acts as a signal that local constraints are less likely to be reliable.

5.3. Limitations and future work

The current study has some limitations. One potential limitation is the use of relatively subtle disruptions to global discourse coherence. This means that our findings are unable to definitively answer how information from the local context is recruited when the global context is completely uninformative. However, our stimuli were designed to contain coherence breaks that are representative of how coherence can be disrupted in natural speech, rather than using artificially meaningless contexts. It could be interesting to investigate how our results compare with less natural disruptions to coherence in future work. But it would be important that these types of trials occur rarely in the study, as there is already evidence that comprehenders are sensitive to the statistical structure of the experiment (Brothers et al., 2017). It is possible that frequent strong coherence violations may lead to a total disengagement in predictive processing.

Another possible suggestion for future work could look at the mechanism behind the processing differences observed for highly predictable critical words, compared to semantically anomalous continuations. One option could be that facilitation occurs for highly predictable words because a correct expectation has been confirmed. However, it is also possible that this processing difference is due to a slowdown in the anomalous condition as expectations have been violated (Kuperberg & Jaeger, 2016; Van Petten & Luka, 2012). In order to adjudicate between these two possibilities, a neutral condition that contains a low cloze, low constraint target sentence would need to be included.

In conclusion, the current study demonstrates that comprehenders are sensitive to relatively subtle changes in the coherence of a discourse and that processing is slowed when these occur. The advantage of a coherent global context emerged at the earliest possible position and extended in time during subsequent processing. Further, our findings demonstrate that a lack of coherence eliminates within-sentence facilitation for predictable words. This suggests that the presence of a narrative shift serves as a cue to temporarily reduce the reliance on predictive processing. Future work could further investigate the nature of this coherence check through the use of finer-grained temporal measures, such as ERPs.

CRediT authorship contribution statement

Georgia-Ann Carter: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Paul Hoffman:** Conceptualization, Writing – review & editing, Supervision.

Declaration of Competing Interest

None.

Data availability

Links to the data and code have been shared within the manuscript.

Acknowledgements

PH was supported by a BBSRC grant (BB/T004444/1). GC was supported in part by the UKRI Centre for Doctoral Training in Natural Language Processing, funded by the UKRI (grant EP/S022481/1) and the University of Edinburgh, School of Informatics and School of Philosophy, Psychology & Language Sciences. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105637.

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1061.

Amsel, B. D., DeLong, K. A., & Kutas, M. (2015). Close, but no garlic: Perceptuomotor and event knowledge activation during language comprehension. *Journal of Memory and Language, 82*, 118–132. https://doi.org/10.1016/j.jml.2015.03.009

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods, 53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv, 1406*, 5823. http://arxiv.org/abs/1406.5823.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657–671. https://doi.org/10.1162/089892999563724

Black, A. (1988). The syntax of conversational coherence. *Discourse Processes, 11*(4), 433–455. https://doi.org/10.1080/01638538809544712

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 15*(3), 607–624. https://doi.org/10.3758/s13415-015-0340-0

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia, 135*, Article 107225. https://doi.org/10.1016/j.neuropsychologia.2019.107225

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language, 93*, 203–216. https://doi.org/10.1016/j.jml.2016.10.002

Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the extra mile: Effects of discourse context on two late Positivities during language comprehension. *Neurobiology of Language, 1*(1), 135–160. https://doi.org/10.1162/nol_a_00006

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language, 98*(2), 159–168. https://doi.org/10.1016/j.bandl.2006.04.005

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language, 56*(1), 103–128. https://doi.org/10.1016/j.jml.2006.07.005

Carter, G., & Nieuwland, M. S. (2022). Predicting Definite and Indefinite Referents During Discourse Comprehension: Evidence from Event-Related Potentials. *Cognitive Science, 46*(2). https://doi.org/10.1111/cogs.13092

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61*, 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016

van Dijk, T. A. (2006). Discourse, context and cognition. *Discourse Studies, 8*(1), 159–177. https://doi.org/10.1177/1461445606059565

Ellis, C., Henderson, A., Wright, H. H., & Rogalski, Y. (2016). Global coherence during discourse production in adults: A review of the literature: Global coherence during discourse production in adults: A review. *International Journal of Language & Communication Disorders, 51*(4), 359–367. https://doi.org/10.1111/1460-6984.12213

Enochson, K., & Culbertson, J. (2015). Collecting psycholinguistic response time data using Amazon mechanical Turk. *PLoS One, 10*(3), Article e0116946. https://doi.org/10.1371/journal.pone.0116946

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Ferreira, F., & Clifton, C. (1986). The Independence of syntactic processing. *Journal of Memory and Language, 25*(3), 348–368.

Fleur, D. S., Flecken, M., Rommers, J., & Nieuwland, M. S. (2020). Definitely saw it coming? The dual nature of the pre-nominal prediction effect. *Cognition, 204*, 104335. https://doi.org/10.1016/j.cognition.2020.104335

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 680.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013). *Word surprisal predicts N400 amplitude during reading. 7*.

Grice, H. P. (1975). Logic and Conversation. In *Speech acts* (pp. 41–58). Brill.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Haeuser, K. I., & Kray, J. (2022). How odd: Diverging effects of predictability and plausibility violations on sentence reading and word memory. *Applied PsychoLinguistics, 43*(5), 1193–1220. https://doi.org/10.1017/S0142716422000364

Hagoort, P. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438–441. https://doi.org/10.1126/science.1095455

Hald, L. A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research, 1146*, 210–218. https://doi.org/10.1016/j.brainres.2007.02.054

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences, 119*(32), Article e2201968119. https://doi.org/10.1073/pnas.2201968119

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General, 124*(1), 62–82. https://doi.org/10.1037/0096-3445.124.1.62

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

Jegerski, J. (2013). Self-paced Reading. In *Research methods in second language psycholinguistics*. Routledge.

Johnson, B. P., Dayan, E., Censor, N., & Cohen, L. G. (2022). Crowdsourcing in cognitive and systems neuroscience. *The Neuroscientist, 28*(5), 425–437. https://doi.org/10.1177/10738584211017018

Just, M. A., & Carpenter, P. A. (1980). A theory of Reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in Reading comprehension. *Journal of Experimental Psychology: General, 111*(2), 228.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kuperberg, G. R. (2021). Tea with Milk? A hierarchical generative framework of sequential event comprehension. *Topics in Cognitive Science, 13*(1), 256–298. https://doi.org/10.1111/tops.12518

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two Positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience, 32*(1), 12–35. https://doi.org/10.1162/jocn_a_01465

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*(2), 99–116. https://doi.org/10.1016/0301-0511(80)90046-0

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161–163. https://doi.org/10.1038/307161a0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13

Ledoux, K., Camblin, C. C., Swaab, T. Y., & Gordon, P. C. (2006). Reading words in discourse: The modulation of lexical priming effects by message-level context. *Behavioral and Cognitive Neuroscience Reviews, 5*(3), 107–127. https://doi.org/10.1177/1534582306289573

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 149–157. https://doi.org/10.1037/0278-7393.16.1.149

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during Reading. *Psychological Science, 14*(6), 648–652. https://doi.org/10.1046/j.0956-7976.2003.psci_1480.x

Medler, D. A., & Binder, J. R. (2005). MCWord: An on-line orthographic database of the English language. http://www.neuro.mcw.edu/mcword/.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language, 66*(4), 545–567. https://doi.org/10.1016/j.jml.2012.01.001

Morris, R. K. (2006). Lexical processing and sentence context effects. In *Handbook of psycholinguistics* (pp. 377–401). Academic Press.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*(2–3), 131–157. https://doi.org/10.1080/01638539809545042

Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 876–886. https://doi.org/10.1037/0278-7393.20.4.876

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., … Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife, 7*, Article e33468. https://doi.org/10.7554/eLife.33468

O'Brien, E. J., & Albrecht, J. E. (1992). *Comprehension strategies in the development of a mental model. 8*.

Olson, D. R., & Filby, N. (1972). On the comprehension of active and passive sentences. *Cognitive Psychology, 3*(3), 361–381. https://doi.org/10.1016/0010-0285(72)90013-8

Peelle, J. E., Miller, R. L., Rogers, C. S., Spehar, B., Sommers, M. S., & Van Engen, K. J. (2020). Completion norms for 3085 English sentence contexts. *Behavior Research Methods, 52*(4), 1795–1799. https://doi.org/10.3758/s13428-020-01351-1

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika, 64*(2), 191–199.

Quante, L., Bölte, J., & Zwitserlood, P. (2018). Dissociating predictability, plausibility and possibility of sentence continuations in reading: Evidence from late-positivity ERPs. *PeerJ, 6*, Article e5717. https://doi.org/10.7717/peerj.5717

R Core Team. (2018). *R foundation for statistical computing; Vienna, Austria. 2014*. https://www.r-project.org/.

Rabovsky, M., & McClelland, J. L. (2020). Quasi-compositional mapping from form to meaning: A neural network-based approach to capturing neural responses during human language comprehension. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 375*(1791), 20190313. https://doi.org/10.1098/rstb.2019.0313

Schustack, W., Ehrlich, S. E., & Rayner, K. (1987). Local and global sources of contextual facilitation. *Journal of Memory and Language, 26*(3), 322–340. https://doi.org/10.1016/0749-596X(87)90117-3

Schwanenflugel, P. J., & White, C. R. (1991). The influence of paragraph information on the processing of upcoming words. *Reading Research Quarterly, 26*(2), 160–177. https://doi.org/10.2307/747980

Stewart, A. J., Kidd, E., & Haigh, M. (2009). Early sensitivity to discourse-level anomalies: Evidence from self-paced Reading. *Discourse Processes, 46*(1), 46–69. https://doi.org/10.1080/01638530802629091

Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. *Journal of Memory and Language, 123*, Article 104311. https://doi.org/10.1016/j.jml.2021.104311

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1266–1282. https://doi.org/10.1037/0278-7393.26.5.1266

Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in Sentence Processing: Intralexical Spreading Activation, Schemas, and Situation Models. *Journal of Psycholinguistic Research, 29*(6), 581–595.

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*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters, 346*(3), 165–168. https://doi.org/10.1016/S0304-3940(03)00599-8

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Willems, R. M., & Peelen, M. V. (2021). How context changes the neural basis of perception and language—ScienceDirect. *Iscience, 24*(5), Article 102392.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage, 62*(1), 356–366. https://doi.org/10.1016/j.neuroimage.2012.04.054

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review, 23*(4), 1028–1034. https://doi.org/10.3758/s13423-015-0864-x

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*(5), 292–297. https://doi.org/10.1111/j.1467-9280.1995.tb00513.x

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 24.