Journal of Memory and Language 137 (2024) 104513

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Eye-movements during reading and noisy-channel inference making

Michael G. Cutter[a,*], Kevin B. Paterson[b,2], Ruth Filik[a,1]

[a] University of Nottingham, UK
[b] University of Leicester, UK

ARTICLE INFO

Keywords:
Eye-movements during reading
Oculomotor control
Sentence processing
Noisy-channel language processing

ABSTRACT

This novel experiment investigates the relationship between readers' eye movements and their use of "noisy channel" inferences when reading implausible sentences, and how this might be affected by cognitive aging. Young (18–26 years) and older (65–87 years) adult participants read sentences which were either plausible or implausible. Crucially, readers could assign a plausible interpretation to the implausible sentences by inferring that a preposition (i.e., *to*) had been unintentionally omitted or included. Our results reveal that readers' fixation locations within such sentences are associated with the likelihood of them inferring the presence or absence of this critical preposition to reach a plausible interpretation. Moreover, our older adults were more likely to make these noisy-channel inferences than the younger adults, potentially because their poorer visual processing and greater linguistic experience promote such inference-making. We propose that the present findings provide novel experimental evidence for a perceptual contribution to noisy-channel inference-making during reading.

Introduction

When reading normally, readers make a series of saccadic eye movements separated by brief fixational pauses on words (each lasting around 250 ms) during which visual information is acquired (for reviews, see Liversedge & Findlay, 2000; Rayner, 1998; 2009). The input to the language processing system therefore comprises a sequence of discrete visual samples of the text. Most saccades are progressive, carrying the reader's gaze forward, although sometimes a reader will also look back at earlier parts of the text, with these 'regressive' saccades accounting for approximately 15 % of saccades by skilled readers of alphabetic scripts like English (e.g., Rayner, 2009). Moreover, not all the words are fixated during reading, and approximately 30 % of words are skipped during the initial (first-pass) processing of a sentence (again by skilled English readers, e.g., Brysbaert et al., 2005). This happens because, while fixating one word, the reader can sometimes also identify the following word in the sentence before they programme their next saccade. As a result, the reader will sometimes then make an eye movement that skips this word to fixate the next word along. When this happens, the skipped word will have been processed outside of central (i.e., foveal) vision, in so-called parafoveal vision, which is of lower retinal acuity (e.g., Balota & Rayner, 1991).

One consequence of this sequence of events in reading is that the perceptual representation that the reader forms of a sentence can be prone to error. This might be due, for instance, to uncertainty about whether a skipped word is actually the word the reader thought it was during parafoveal processing, or whether a word is actually in the sentence location that the reader encoded it as being. We focus on this issue with the present research, by considering how the nature of the perceptual input to the sentence comprehension system can influence the probability of a reader misinterpreting a sentence, as well as by considering how readers might alter their visual sampling strategy in response to encountering implausible material. The work we present offers novel empirical evidence of theoretical importance by linking eye-movement behavior during reading to predictions derived from the noisy-channel theory of language processing (Gibson et al., 2013; Gibson et al., 2017; Ryskin et al., 2018; Ryskin et al., 2021; Zhang et al., 2023). This is an influential theory of language processing that has been largely developed outside of the field of eye-movement research (although see Levy et al., 2009).

The type of sentences we examine have previously been investigated by Gibson et al., (2013; see also Gibson et al., 2017; Ryskin et al., 2018) in the context of the noisy-channel theory. According to the theory, the language comprehension system takes account of uncertainty in the

* Corresponding author at: School of Psychology, University of Nottingham, University Park Campus, Nottingham NG7 2RD, UK.
E-mail address: Michael_cutter@hotmail.com (M.G. Cutter).
[1] School of Psychology, University of Nottingham, Nottingham, UK.
[2] School of Psychology and Vision Sciences, University of Leicester, Leicester, UK.

https://doi.org/10.1016/j.jml.2024.104513
Received 26 February 2023; Received in revised form 15 February 2024; Accepted 16 February 2024
Available online 28 February 2024

nature of perceptual input. In particular, the theory proposes that the language comprehension system assumes that perceptual input will contain errors and therefore that the reader may have to modify this input (e.g., by adding, deleting or re-ordering words) to achieve a plausible interpretation of a sentence. With the present research, we examined the processing of sentences which can be made plausible or implausible via the presence or absence of the word *to*, as below:

1. The mother handed the daughter the candle.
2. The mother handed the daughter to the candle.
3. The mother handed the candle to the daughter.
4. The mother handed the candle the daughter.

Sentence 1 is an example of a double-object (DO) construction that describes a plausible event (i.e., the action of handing a daughter a candle fits with our world knowledge). However, inserting the word *to* (as in Sentence 2) transforms this sentence into an implausible prepositional phrase object (PO) construction, where the described event (handing a human to a candle) makes little sense. Sentence 3 describes the same plausible event as Sentence 1 using a PO structure, with the order of *daughter* and *candle* reversed. Sentence 4 is again implausible, due to the removal of the word *to* from Sentence 3 creating a DO structure in which a human is again handed to a candle. Gibson et al. considered whether readers are likely to revise the structure of these implausible sentences to achieve a plausible interpretation (e.g., by mentally adding or deleting the preposition *to*). Participants in their experiment read sentences like Examples 1–4, and their comprehension was tested using questions such as "Did the candle receive something/someone?". The answer to this question should be "No" for Sentences 1 and 3, as these sentences describe an event in which it was the daughter that received something. Gibson et al. found that participants were highly accurate at answering questions following sentences like these, achieving near ceiling performance in this task. For Sentences 2 and 4, the correct answer should be "Yes". The accuracy of participants' responses for these sentences was much lower, however, and barely above chance (~55 % accuracy). Gibson et al. took this as evidence that readers were systematically misinterpreting these sentences as their more plausible alternative. Such misinterpretations or misrepresentations of the sentences are also observed when comprehension is probed via tasks in which participants either transcribe the sentence while being told to explicitly correct any errors (Ryskin et al., 2018), match the sentence with pictures illustrating the alternative interpretations (Warren et al., 2017), or act out the sentence using stick puppets (Gibson et al., 2015).

Gibson et al. (2013) explained these effects in terms of noisy-channel inferences (see also Levy, 2008). As a starting point, they assume that the communication and perception of a stimulus is a fundamentally noisy process, in which information is produced and sampled in such a manner that the resulting percept is unlikely to perfectly recreate the stimulus as intended by the producer. Following this approach, it is assumed that the language comprehension system takes account of this possibility by assessing how likely the sentence perceived by the reader (or listener) was the sentence intended by the writer (or speaker). In cases where what is perceived by the reader seems unlikely to be the sentence intended by the writer (e.g., when a sentence is nonsensical), the system calculates whether a more likely meaning could be arrived at by assuming a particular error. For examples like Sentences 1 and 3 above, no such inferences are required, as the sentences already are perfectly sensible. For Sentences 2 and 4, however, such inferences are made, due to the implausible nature of the message being communicated. This allows readers to arrive at a sensible meaning for Sentence 2 by assuming that either they had somehow 'imagined' the presence of the word *to* or the producer had accidentally included this within the utterance. A sensible meaning for Sentence 4 can be achieved by readers assuming that the word *to* was supposed to be present after *candle*, and that they somehow missed this word or the sentence producer somehow failed to include it. Note that in Gibson et al.'s framework, this latter type of error is more likely to occur due to the Bayesian size principle (see MacKay, 2003), as a word missing from a sentence is considered more likely than for people to have selected a specific word (i.e., *to*) from their mental lexicon and imagined it or accidentally inserted it within a sentence. A final aspect of Gibson et al.'s account is that people are more likely to make misinterpretations when the transformations they could make to the sentence are relatively small as opposed to relatively large. For example, Gibson et al. argue that people make many errors for the example sentences shown above in which the transformation involves the presence/absence of *to*. However, people make fewer errors for implausible active sentences such as *The ball kicked the girl*, for which they would need to assume that the words *was* and *by* both were erroneously missing from the sentence after *ball* and *kicked*, respectively.

Noisy-channel accounts incorporate the assumption that the internal noise model can vary across individuals. For example, Futrell and Gibson (2016) argued that non-native comprehenders will rely more heavily on their real-world knowledge than knowledge about syntactic structure because of their more impoverished knowledge of syntax in their second language. Similarly, Warren et al., (2017; see also Gibson et al., 2015) observed a greater dependence upon the prior probability of an interpretation in participants with aphasia than age matched controls. Comprehenders also appear able to adjust their noise models to the current situation. For example, Gibson et al. (2017) found that comprehenders are more likely to interpret implausible utterances as their plausible alternatives when spoken in a foreign accent, based on the assumption that non-native speakers are more likely to make a production error. Gibson et al. (2013) also found that readers are more likely to assign a non-literal interpretation to sentences when filler items presented alongside these stimuli suggest that the probability of a syntactic error is high. By contrast, readers are less likely to assign a non-literal interpretation when the characteristics of the filler items suggest that the likelihood of something implausible genuinely being communicated is high (see Ryskin et al., 2018). Altogether, this suggests that the probability of comprehenders making noisy-channel inferences is modulated by both internal and external factors, and in particular how strongly readers believe an error may have been made in sentence production.

Most the work discussed above has been interpreted in terms of participants making noisy-channel inferences because they assume that an error has been made in sentence production, as opposed to being a result of the participant misperceiving the sentence. Indeed, a recent study (Liu et al., 2021) suggests that noisy-channel inferences occur purely due to readers assuming an error in production. Here, participants answered comprehension questions about implausible sentences, and then were asked to directly re-produce what they had read. Despite obtaining standard misinterpretation effects, Liu et al. found that participants were 99 % accurate when reproducing the sentences. Liu et al. took this as evidence that noisy-channel inferences are almost entirely uninfluenced by participants' perceptions of the sentences. With the present research, we further test whether perceptual factors can influence the probability of a reader making a noisy-channel inference by focusing on two factors that might modulate such effects. Specifically, we examine how the pattern of eye-movements during sentence reading may influence the probability of people inferring a non-literal meaning by affecting the amount of perceptual evidence they have for the presence/absence of a word. In addition, we examine whether perceptual and cognitive changes associated with cognitive aging increase the likelihood of a reader making a noisy-channel inference, for reasons discussed below.

One source of perceptual error that may affect the probability of readers making a noisy-channel inference is in the acquisition of perceptual samples via the oculomotor and visual system. As discussed above, readers typically do not fixate every word in a sentence. Moreover, even when they do, words within the sentence may not be sampled in serial order, due both to readers' frequently skipping words and making regressive eye-movements to reinspect earlier words. Consider sentences like *The mother handed the candle the daughter.* If a reader does not fixate close to the gap between the words *candle* and *the*, they may be more likely to infer that they missed the word *to* in this location, and so misinterpret the sentence. Conversely, for sentences like *The mother handed the daughter to the candle*, readers may be more likely to misinterpret the sentence by omitting the word *to* when they do not fixate this word; that is, in cases where they skip the word and do not fixate it in any subsequent re-reading. By examining patterns of eye-movements across the sentences used by Gibson et al. and seeing how this links to responses to comprehension probes, we aim to determine whether there is a link between eye-movement behavior and use of noisy-channel inferences, and more generally whether perceptual factors influence misinterpretation rates.

A further issue concerns how the implausibility of these sentences might in turn affect readers' visual sampling strategies. Of specific relevance is research by Levy et al. (2009) examining whether readers maintain and act upon word-level uncertainty during reading. In line with noisy-channel accounts, Levy et al. proposed that because the input to the sentence processor is essentially noisy and therefore unreliable, readers may maintain uncertainty about the identity of the words they encounter, and then interactively recruit grammatical analysis and subsequent perceptual input to resolve uncertainty. For example, upon encountering the preposition *at* in a sentence, readers may assign a probability of 0.85 that this word is indeed *at*, but also assign a probability of 0.10 and 0.05 that the word is actually either *as* or *and*, respectively. Levy et al. tested this possibility by constructing sentence stimuli that included the word *at* but where subsequent linguistic information suggested that either the word *as* or *and* would be more appropriate in the syntactic structure being constructed. They assumed that readers should be more likely to re-read earlier parts of the sentence and experience more processing difficulty under these circumstances as a direct consequence of their uncertainty about the preceding text. Thus, evidence for such processing difficulty within the eye-movement record would represent important evidence for noisy-channel accounts of language processing. Although Levy et al. (2009) reported evidence in favour of this possibility, a recent pre-registered replication study by Cutter et al. (2022b) with considerably more statistical power could not replicate this finding, with a Bayesian analysis finding evidence against the key effect. Furthermore, other recent studies report evidence against the idea that readers maintain and act upon uncertainty about prior information in reading (Cutter et al., 2022a; Paape et al., 2022). Given this evidence against the word-level uncertainty aspect of noisy-channel processing, it is vital to determine whether readers maintain and act upon other forms of uncertainty during reading relating to the presence/absence of words, and whether these effects are more likely to manifest when a sentence is implausible as opposed to simply being of a low probability syntactic structure as in prior work.

A final issue of interest concerned how effects of cognitive aging might affect readers' propensity to make noisy-channel inferences. As people age, they experience sensory and cognitive declines (see Salthouse, 2010). This includes declines in visual functioning (see Owsley, 2011, for a review), which may lead to a greater deal of noise and uncertainty in any perceptual representation formed of a written sentence (Cutter et al., 2022a). We might therefore expect older adults to make more noisy-channel inferences simply because of declines in visual acuity. In addition, as people age, they tend to acquire greater linguistic knowledge (Hartshorne & Germine, 2015) and more general world knowledge (Horn & Cattell, 1967), which could increase the influence of prior knowledge in noisy-channel inferences (see Christianson et al., 2006 for a similar argument in relation to aging and good-enough processing).

Also of relevance here is the argument that patterns of eye-movements during reading suggest that older adults exhibit a different pattern of fixations and saccades across a sentence (McGowan & Reichle, 2018; Rayner et al., 2006; see also Zhang et al., 2022). This is based on the observation that, compared to young adults, older adults tend to make both longer forward eye movements and more regressions back to earlier points in a text. Putting aside what causes this pattern of eye-movements, it should be clear that such differences might increase noise in the perception of language. Consider again the example sentence, *The mother handed the candle the daughter*. As already discussed, this sentence can be normalised via the reader assuming that they failed to perceive the word *to* after *candle*. Crucially, the likelihood of this inference may be influenced by eye movement behavior and, in particular, may be more likely where readers make longer forward saccades and so are generally more likely to fail to notice words in the text.

Other recent work has attempted to link eye-movement behavior to the failure to notice errors in text (see Huang & Staub, 2021a, for a recent theoretical review). For example, Staub et al. (2019) tracked participants' eye movements while they read sentences that were either grammatical, erroneously featured the article *the* twice in succession, erroneously featured a repeated noun, or erroneously omitted the article *the*. Once they finished reading each sentence, participants were asked whether there was anything wrong with it. Staub et al. found that readers generally responded correctly following a grammatical sentence or for sentences with a repeated noun, but regularly responded incorrectly to sentences which omitted the article *the* (33 % of the time) or repeated the article *the* (54 % of the time). Examination of the eye-movement data revealed an effect of fixation location on the likely detection of an error, such that detection rates were elevated if participants fixated the repeated articles (i.e., 66 % correct responses). For sentences in which *the* had been erroneously omitted, fixation of the words flanking where *the* should have appeared had modest effects on responses, such that the omission was detected on 71 % of trials in which both the preceding and following words were fixated and on 64 % of trials where only one of these words was fixated. In another study, Huang and Staub (2021b) examined the link between eye-movement behavior and the likelihood of readers correctly classifying a sentence containing two transposed words as being ungrammatical (e.g. *The white was dog big*). They observed that readers were more likely to report this ungrammaticality when they had fixated both transposed words (e.g., *dog* and *was*) than when they had not. In addition, they found only measurable disruption in reading behavior due to a word transposition when readers judged the sentence to be ungrammatical. This suggested that, on trials in which participants made an error, they failed to detect the transposition. Thus, this work suggests a role for the way readers visually sample text in determining whether they detect errors or ungrammaticalities within a sentence. The present work builds upon these findings by extending this link to the extraction of meaning from a sentence, as opposed to explicit judgements of ungrammaticality. Establishing the existence (or lack thereof) of such an effect for these types of sentences seems particularly important in light of Liu et al.'s (2021) claim that readers are making noisy-channel inferences exclusively due to assuming producer-level errors.

The present study

With the present study, we investigated 1) how eye-movement behavior influences sentence interpretation; 2) how violations of plausibility influence eye-movements when a more plausible meaning can be reached via the insertion or deletion of *to*; and 3) how aging affects noisy-channel inferences and whether any such age differences can be attributed to differences in eye-movement behavior. We investigated these issues by asking a group of young adults (aged 18–26 years) and a group of older adults (aged 65+ years) to read sentences similar to those used by Gibson et al. (2013) while their eye movements were recorded. In our experiment, participants read a brief context sentence (e.g., *It was dark due to the power cut.*) followed by a sentence that was either plausible or implausible as a DO sentence (e.g., *The mother gave the daughter the candle*; *The mother gave the candle the daughter*) or PO sentence (e.g., *The mother gave the candle to the daughter*; *The mother gave the daughter to* *the candle*). These sentence-pairs were followed by comprehension questions of a type typically used to assess noisy-channel processing (e.g., *Did the candle receive something/someone?*), in which an incorrect answer following implausible sentences would suggest that the participant had made a noisy-channel inference. In addition to collecting eye-movement and comprehension data for sentences, we assessed the participants' visual and cognitive capabilities. As discussed above, declines in visuo-perceptual abilities and increases in verbal knowledge might lead older adults to make a greater number of noisy-channel inferences. Accordingly, it was important that we first establish whether our participants showed the anticipated age differences in these visual and cognitive capabilities.

We considered hypotheses about how the above factors might affect noisy-channel inferences. First, we hypothesised that older adults would make more noisy-channel inferences than young adults. There were several reasons for this prediction. As discussed above, it is likely that visual declines associated with aging (Owsley, 2011) increase the level of perceptual noise that participants must take account of during reading. Furthermore, the increase that is typically observed in older adults' verbal knowledge may cause them to rely more on world-knowledge, providing them with stronger prior expectations of what is a sensible utterance. Increases in linguistic ability may also result in older adults being more likely to derive an alternative syntactic structure. Finally, older adults may exhibit a different pattern of eye-movement behavior, compared to younger adults, for the sentences used in the current study, as indicated by prior research (e.g., see Rayner et al., 2006; Zhang et al., 2022). While two prior studies have tested older adults' propensity to make noisy-channel inferences, these primarily were aimed at assessing effects of aphasia on noisy-channel inference making, with one study examining only a small sample of non-aphasic older adults (i.e., n = 7; Gibson et al., 2015) while the other did not compare its older adult control group with a young adult group (Warren et al., 2017). Accordingly, the specific effects of cognitive aging on noisy-channel inference-making require further investigation.

As discussed above, we were also interested in examining how eye-movement behavior on individual trials might influence the probability of a noisy-channel inference. In terms of the implausible DO sentences (e.g., *The mother handed the candle the daughter*), we hypothesised that participants would be more likely to answer comprehension questions correctly when they fixated the final character of *candle*, the first character of *the* preceding daughter, or the space in-between (i.e., the location where *to* would have appeared).[3] For the implausible PO sentences (e.g., *The mother handed the daughter to the candle*), we hypothesised that participants would be more likely to answer comprehension questions correctly if they fixated the word *to*. In both cases, linguistic input to the sentence processing system should be assessed as being more reliable when a relatively high level of perceptual evidence is obtained from the sentence location at which a word insertion/omission might have occurred. By comparison, we hypothesised that eye-movement behavior in other less perceptually relevant areas of the sentence may have less of an effect on misinterpretation rates. We addressed this hypothesis by selecting a further region in each sentence type which contained a similar type and amount of perceptual and linguistic information as our target region, but where word insertion or deletion would not affect sentence meaning.

It is also interesting to consider how differences in the eye movement behaviour of the two age groups might affect how they interpret our sentences. As already noted, visual declines in older age might lead older readers to make more noisy-channel inferences. If so, we can imagine two ways in which this might impact on reading. First, it might simply be that older adults maintain more uncertainty than young adults about the totality of the perceptual evidence they obtain. Another possibility is that older adults are more uncertain about each piece of perceptual evidence they obtain. If this latter possibility is the case, then young readers might obtain a larger benefit from each new piece of evidence compared to the older adults. We might therefore see an increase in the misinterpretation rate for older compared to young adults as the two groups acquire more perceptual evidence from the target region. This effect might be observed most clearly in continuous measures such as fixation count and the total time spent in at the location in which an insertion or deletion could be made to reach a plausible interpretation of an implausible sentences.

A final question concerns how sentence type affects eye-movement behavior. First is the issue of whether readers tend to obtain more perceptual evidence from the site of a potential insertion/omission in implausible sentences. Such a tendency would suggest that, when encountering a sentence of low semantic probability, readers bolster their perceptual representation, including by making more regressions to re-read parts of the text, to determine whether they had made a perceptual encoding error during first-pass reading (see also Levy et al., 2009, and Sturt & Kwon, 2018, for evidence that readers extract additional perceptual information from words during regressions). A further issue concerns whether this varies systematically across age groups. Recall that prior research shows that older adults tend to skip more words in a sentence while also being more likely to make regressions. One possibility is that older adults extract a different amount of information from our sentences compared to young adults. It is therefore important to determine if this is the case, as this could obscure any age group effects in misinterpretation rates.

Finally, it is worth considering to what extent the time spent reading our sentences is driven by whether the readers make a noisy-channel inference. As mentioned above, Huang and Staub (2021b) found that eye movement behavior was not disrupted by word transpositions when the readers did not report that the sentence was ungrammatical. One possibility is that we might observe a similar effect for the implausible sentences in our experiment, such that there is less disruption to eye movement behavior when readers assign a plausible meaning to these sentences. This might indicate that readers can make a noisy-channel inference sufficiently rapidly that the implausibility is not detected. If so, sentence reading times might reveal an interaction between sentence plausibility and whether participants made a noisy-channel inference, such that reading times do not differ for implausible sentences when a noisy-channel inference is made, as compared to the plausible sentences. By comparison, we would expect to observe standard plausibility effects when a noisy-channel inference is not made to correct the meaning of the implausible sentences.

Taken together, answers to these questions have the potential to further our understanding of how perceptual information acquired during reading can affect the use of noisy-channel inferences to interpret implausible sentences. Our work might also shed light on whether readers respond to implausible input by resampling visual information that might allow the reader to confirm its veracity. We consider this an important contribution to our understanding of the oculomotor control system in general, and of the highly impactful noisy-channel approach to sentence processing, which has been underexplored using eye movement methods.

Data availability

A registered, permanent version of our data, materials, and analysis code can be found at https://osf.io/e6nyf.

[3] We choose this three-character region as a compromise between where *to* could have appeared in the sentence (i.e., directly in the space between two words) and a larger region which would feasibly provide us with enough eye-movement data to perform our analyses.

Table 1
Visual and Cognitive Assessment of Young and Older Adult Participants.

	Young Adults				Older Adults			
	M	SD	Min	Max	M	SD	Min	Max
LogMar (40 cm)	0.01	0.10	0.10	−0.10	0.26	0.14	0.4	−0.1
LogMar (97 cm)	0.32	0.07	0.20	0.40	0.50	0.10	0.7	0.3
Contrast-Acuity	1.82	0.11	1.65	1.95	1.67	0.15	1.35	1.65
Working Memory	21.94	4.50	17	31	21.41	3.99	15	33
Verbal Comprehension	39.15	6.74	24	47	44.89	5.43	31	55
Processing Speed	22.46	4.31	17	38	17.67	3.77	9	30

Method

Participants

Participants were 48 young adults (mean age = 20.1 years; age range = 18–26 years; 30 female, 18 male; mean years of education 15.1; mean hours reading per week = 15) and 52 older adults (mean age = 71.1 years; age range = 65–87 years; 39 female, 13 male; mean years of education = 16.4; mean hours reading per week = 11), all of whom were native speakers of English from the Nottingham area. The older adults were screened for unimpaired cognitive abilities using the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) and were included in analyses if they scored at least 26 (out of 30). All participants scored 26 or better (M = 28, SD = 1.39, range = 26–30). Participants were compensated for taking part in the study with research credits (if they were first year Psychology students) or an inconvenience allowance of £10 per hour.

We assessed the visual and cognitive capabilities of the two participant groups. The results are shown in Table 1. In terms of visual abilities, acuity was tested using a logMar chart (Ferris & Bailey, 1996) at both a 40 cm distance and the reading distance (i.e., 97 cm) in the experiment. Contrast sensitivity was assessed using a Pelli-Robson contrast-sensitivity chart (see Elliott et al., 1990) at 2 m. The acuity test produced typical age group differences, while no participants exhibited clinically low vision (i.e., all scores > 0.5 logMar). Moreover, even the worst performing older adult could perceive text of the size used in the experiment at the 97 cm test distance. The older adults had poorer contrast sensitivity compared to the young adults, as is typical. Unpaired *t*-tests demonstrated that the two age groups differed significantly on all three visual tests (all *ps* < 0.001). This confirmed that the older adult group had poorer visual capabilities than the young adult group.

Cognitive abilities were assessed using working memory, processing speed, and verbal comprehension sub-tests of the WAIS-IV (Wechsler, 2008). Note that the scoring guide for the subtests provides a scaled score for different age groups, such that participants are assessed relative to their age-group peers. However, for research on cognitive aging, it makes sense to scale participants of both age groups in the same way, to compare performance across age groups. Accordingly, the scores in Table 1 are scaled scores for a 20 to 25 year-old for both groups. As expected, the older adults scored higher on verbal comprehension than the young adults ($t(98) = -5.22$, $p < 0.001$), as is typical (Kaufman et al., 2016). The older adults scored lower in terms of processing speed ($t(98) = 5.84$, $p < 0.001$), while the two age groups produced similar scores for working memory ($t(98) = 0.54$, $p = 0.594$).

Design & stimuli

Each participant read 40 experimental stimuli in which a context sentence was followed by a plausible or implausible sentence with either a double-object or prepositional phrase object structure. This context sentence was generally supportive of the plausible meaning of the item, and identical across conditions. Each stimulus was followed by a yes/no comprehension question targeted towards assessing whether participants interpreted the sentence literally, or non-literally via noisy-channel inference, following the approach taken in previous studies of noisy-channel inference-making. An example item, in all four conditions, was as follows:

Context: It was dark due to the power cut.
Double-Object Plausible: The mother handed the daughter the candle.
Double-Object Implausible: The mother handed the candle the daughter.
Prepositional Object Plausible: The mother handed the candle to the daughter.
Prepositional Object Implausible: The mother handed the daughter to the candle.
Question: Did the daughter receive something/someone?

For half of these items a “yes” answer indicates a literal interpretation, while for the other half a “no” answer indicates a literal interpretation. Twenty of the stimuli were from previous studies of noisy-channel inferences in plausible and implausible PO and DO sentences, and the remaining 20 items were newly created. Our stimuli are available at https://osf.io/qa2fe. The within-participants element of the experiment had a 2 (Plausible vs. Implausible) × 2 (DO vs. PO) design. However, the statistical models used to answer some of our theoretical questions will focus on sub-sets of these conditions, as discussed in detail in the Results section. In addition to our within-participants variables, we examined the between-participant variable of age group as a categorical variable (Young vs. Older).

The experimental stimuli were intermingled with 74 filler items. Thirty-eight of these were a part of an experiment examining implicit causality. The other 36 items were part of an experiment examining the processing of filler-gap structures. Half of these items included a filler-gap structure; the other half included a prepositional phrase structure. Twenty-four of these filler items were followed by comprehension questions, with these again having an equal balance of yes/no answers.

Apparatus and procedure

Participants’ eye movements were recorded using a desktop-mounted SR Research EyeLink 1000 running at 1000 Hz. Sentence stimuli were shown as a single line of text in size 23 Courier New font using a BenQ XL2430 24″ monitor, with black text on a grey background. Viewing distance was 97 cm, with ~ 3.6 characters subtending approximately 1 degree of visual angle.

Participants were informed about the study and gave written consent on arrival to the lab, followed by an assessment of their visual abilities assessed using the LogMar and Pelli-Robson chart. Participants were then set up on the eye-tracker, using a chin and head rest to minimise head movements. The eye-tracker was calibrated to their eye movements using a three-point horizontal calibration grid. If during a validation procedure a participant had an average error greater than .50 degrees or any individual errors above .50 degrees, the system was re-calibrated. During the experiment, each trial consisted of 1) a drift check at the center of the screen, 2) a drift check at the left of the screen, in the same position as the first character of the sentence, 3) a gaze

The mother handed the daughter the candle.
The mother handed the candle the daughter.
The mother handed the candle to the daughter.
The mother handed the daughter to the candle.

Fig. 1. An example of the interest areas used for data analysis in our study. The perceptually relevant region, in which a word could be inserted/deleted to change the sentence meaning, is shown in the solid box. The control region is shown in the dashed box.

Table 2
Mean (standard error) Comprehension Accuracy.

	Overall		Young Adults		Older Adults	
	Plausible	Implausible	Plausible	Implausible	Plausible	Implausible
DO	.90 (0.02)	.59 (0.02)	.90 (0.01)	.65 (0.02)	.90 (0.01)	.53 (0.02)
PO	.91 (0.01)	.81 (0.01)	.89 (0.01)	.84 (0.02)	.92 (0.01)	.77 (0.02)

contingent box in the same position as the second drift check, 4) the experimental sentence itself and 5) a comprehension question on certain trials. Participants were recalibrated if they returned an error above .50 on either drift check on two consecutive trials.

Participants were asked to read silently for comprehension. To minimize blinking during the reading of the sentences, participants were asked to deliberately blink when either of the drift checks or the comprehension question was on the screen. They were informed that the yes–no comprehension question following the sentences might sometimes seem a bit odd, but to try their best to answer on the basis of what they had just read. Questions were answered using a computer mouse, with the left button used for "yes" responses and the right button for "no" responses. Participants were given the opportunity to take breaks during the experiment. At the end of the experiment, participants were asked if they noticed anything odd or unusual about the task, stimuli, or questions, and were debriefed as to the purpose of the experiment.

After the eye-tracking experiment participants completed the Montreal Cognitive Assessment and the relevant subsets of the WAIS-IV.

Results

After completing data collection, we used SR Research DataViewer to prepare and export our data for analysis. Following standard procedures, we removed fixations longer than 800 ms, merged fixations below 80 ms with any fixations less than 0.5 degrees away, merged fixations below 40 ms with fixations less than 1.25 degrees away, and finally removed any remaining fixations that lasted less than 80 ms. We also manually cleaned the data via playback of the recorded data within DataViewer to determine if any trials included substantial tracker loss. This accounted for a total of 78 experimental items (1.95 % of data), which were removed.

We constructed interest areas using DataViewer. For each sentence, we included a perceptually relevant interest area and a control interest area. The perceptually relevant area consisted of the sentence region where insertion/deletion of *to* could transform sentence meaning, while the control region consisted of a similar area in terms of content but where edits would not lead to an alternative interpretation of the sentence. These interest areas are illustrated in Fig. 1. For the double-object sentences, our perceptually relevant interest area (shown as solid outlines) consisted of the final character of the first object (e.g., the *e* in *candle* for the implausible sentence; the *r* in *daughter* for the plausible sentence), the space after this word, and the *t* in the following article *the*. For the prepositional object sentences, the perceptually relevant interest area consisted of the word *to* as well as half of each space preceding and following this word. The corresponding control regions are shown for each sentence structure using dashed outlines. For the double-object sentences, this comprised the final letter of the verb (e.g., *d* in *handed*), the space following this, and the first letter of the following article *the*. We chose this as the control area as it closely matched the characteristics of the perceptually relevant region, by comprising the end of a content word, a space, and the start of a function word. For the prepositional object sentences, the control area consisted of the word *the* immediately following the verb, with no spaces to either side included in the interest area. This region was chosen, because much like *to,* it is a function word. The space either side of this control region was excluded to ensure that the control area was spatially equivalent to its matched perceptually relevant area.

Formal data analyses were conducted in R (R Core Team, 2020). All analysis scripts and data files are available at https://osf.io/e6nyf. We used the brms package (Bürkner, 2018) to construct Bayesian mixed-effect regression models. Our models used weakly informative priors of *Normal*(0, 10) for the model intercept and *Normal*(0, 1) for each predictor effect, with a regularization of 2 on the covariance matrix of random effects. These priors would be considered weakly informative and exert minimal influence upon model predictions. Models were run initially with four chains of 8000 iterations each, with the first 1000 iterations used as warm-up. Additional iterations were added in cases where the output of a model suggested a convergence issue. For models with a binomial outcome variable, the model family was set to Bernoulli; for those with a continuous outcome variable, the model family was set to lognormal; and for models with a count outcome the model family was set to Poisson. All models used a maximal random effects structure. For each model, we report a median effect estimate (i.e., the most likely size for the effect), the upper and lower bound for a 95 % credible interval (i.e., the values between which there is a 95 % probability the parameter value lies) and $p(\widehat{b} > 0)$ or $p(\widehat{b} < 0)$ (i.e., the probability that the variable's effect was either positive or negative, respectively). Generally, we consider there to be evidence for an effect of a variable of interest if the credible interval does not include values of 0 and $p(\widehat{b} >/< 0)$ is 0.975 or higher, as in these circumstances we can be relatively sure that an effect in the opposite direction is not credible (see Kruschke et al., 2012; Nicenboim & Vasishth, 2016). To be clear, $p(\widehat{b} >/< 0)$ indicates the probability that the effect is in a particular direction assuming it exists, rather than the probability that an effect exists at all. Contrasts were coded to examine main effects, such that one level of each categorical variable was coded as 0.5 and the other coded as −0.5.

Table 3
Mean (Standard Error) Eye-Movement Measures for Each Interest Area.

	DO				PO			
	Plausible		Implausible		Plausible		Implausible	
	PR	Control	PR	Control	PR	Control	PR	Control
Skipping	0.63 (0.02)	0.66 (0.02)	0.64 (0.02)	0.65 (0.02)	0.73 (0.01)	0.73 (0.01)	0.72 (0.01)	0.73 (0.01)
Fix. Prob.	0.60 (0.02)	0.66 (0.02)	0.75 (0.01)	0.72 (0.01)	0.61 (0.02)	0.56 (0.02)	0.77 (0.01)	0.65 (0.02)
Fix. Count	0.98 (0.03)	1.02 (0.03)	1.51 (0.04)	1.26 (0.04)	0.91 (0.03)	0.82 (0.03)	1.36 (0.04)	1.06 (0.03)
Total Time	395 (11)	352 (8)	493 (12)	408 (10)	366 (10)	316 (8)	465 (12)	371 (9)

For some of our more complex models, we are selective about which effects are reported in the main text of our manuscript, focusing on effects relevant to our hypotheses; tables with full model output can be found in the Appendix.

We first examined whether we could replicate effects observed by Gibson et al. (2013) for these sentence types. This would entail comprehension being at near ceiling for plausible sentences presented in a DO or PO syntactic structure, comprehension being lower for implausible PO structures, and lower still for implausible DO structures. In other words, we expected an interaction between sentence plausibility and sentence structure. It is also at this point that we first address the question of whether older adults are more likely to make noisy-channel inferences than young adults. To address these questions, we fitted a Bayesian mixed-effects regression model with comprehension accuracy as an outcome variable, and sentence plausibility, sentence structure, and participant's age group, and three- and two-way interactions between these effects as predictor variables. We also included an effect of the animacy of the receiver in our comprehension questions (i.e., in *did the girl receive something/someone* the receiver is animate, whereas in *did the car receive something/someone* the receiver is inanimate), with this interacting with all other variables. This factor was included in our analysis to establish whether there was any influence of this between-item variable on any age-group effects.

Mean comprehension scores are shown in Table 2. As these show, the experiment replicated Gibson et al.'s standard effects, with an effect of plausibility ($b = 2.42$, CrI[1.83, 3.02], $p(\widehat{b} > 0) = 1$) and of sentence structure ($b = 1.33$, CrI[0.89, 1.79], $p(\widehat{b} > 0) = 1$), and an interaction between plausibility and sentence structure ($b = -0.80$, CrI[−1.48, −0.13], $p(\widehat{b} < 0) = 0.990$). Note that the overall effect is smaller than typically observed in prior studies, especially for the prepositional object sentences, which produced a small difference between plausible and implausible sentences. Turning to age-group effects, we found a reliable interaction between age group and plausibility ($b = 0.97$, CrI[0.15, 1.79], $p(\widehat{b} > 0) = 0.990$). As can be seen in Table 2, the older adults were more strongly affected by plausibility than young adults. This effect took the pattern of both age groups performing at near ceiling for the plausible sentences, with older adults experiencing a larger drop in accuracy for the implausible conditions than the young adults. This is the key finding from this analysis. We also confirmed that both age groups more were likely to make a noisy-channel inference when the referent was animate as opposed to inanimate, such that animacy interacted with plausibility ($b = 1.89$, CrI[0.87, 2.88], $p(\widehat{b} > 0) = 0.999$), but with no other reliable effects involving animacy.

Eye-movements as a function of sentence type

We now turn to the question of how sentence type affected eye-movement behavior. To do so, we focus on four measures of eye-movement behavior that we considered relevant to whether readers make noisy-channel inferences. These are the probability of a region being skipped during first-pass reading, the probability of a region being fixated at all regardless of whether this occurred during first-pass or subsequent re-reading of the region (i.e., including cases in which the region is skipped and then subsequently fixated), the number of fixations each region received before participants completed reading, and the total time for which a region was fixated before participants completed reading. For total time, trials in which either interest area was never fixated (i.e., a dwell time of 0 ms) were removed from analyses. This decision was taken as leaving these instances in would create a highly unusual distribution of this variable. For both total time and fixation count, we excluded values more than 4 standard deviations from the grand mean. This was computed separately for the perceptually relevant region and the control region, as we a priori expected a difference between these two distributions. For total time this accounted for 0.89 % of data in the perceptually relevant region and 0.59 % of data in the control region. For fixation count, it accounted for 0.59 % of data in the perceptually relevant region and 0.48 % of data in the control region.

For each measure, we constructed a Bayesian mixed-effects regression model in which sentence plausibility (plausible vs. implausible), sentence structure (double-object vs. prepositional object), interest area (perceptually relevant vs. control region), and age group (young adults vs. older adults) were treated as predictor effects. For these models, we included main effects of plausibility, interest area, age group, and interactions between these three factors, with the effects nested under sentence structure. We took the approach of building nested effects models, as interactions between sentence structure and other factors were not of theoretical interest. Therefore, constructing models with additional two-, three-, and four-way interactions made little sense. We were more interested in accounting for variance due to sentence structure when estimating effects of interest (i.e., plausibility, interest area, and age group; see Schad et al., 2020 for a discussion of contrast coding and the use of nested effects models) and establishing eye-movement effects in two separate sentence types (double-object and prepositional phrase object). Conditional means collapsed across age groups are shown in Table 3. To re-iterate, our hypothesis was that readers would spend more time reading the implausible sentences compared to the plausible sentences, with more of this reading behaviour being focused on the perceptually relevant region compared to the control region. Thus, the key effect here was the interaction between plausibility and interest area.

In the model for word skipping (i.e., the probability of a region being skipped during first-pass reading regardless of whether it was then fixated in subsequent re-reading), there were no interactions between plausibility and interest area for either DO ($b = -0.01$, CrI[−0.29, 0.27], $p(\widehat{b} < 0) = 0.531$) or PO ($b = 0.10$, CrI[−0.24, 0.42], $p(\widehat{b} > 0) = 0.724$) sentences. There was however evidence for an effect of age group in DO sentences ($b = 0.29$, CrI[0.03, 0.55], $p(\widehat{b} > 0) = 0.986$) with older adults skipping both regions more than young adults.

The model for fixation probability (i.e., the probability of a region being fixated at all regardless of whether this occurred during first-pass or subsequent re-reading after a later region was fixated) revealed an interaction between plausibility and interest area for both DO ($b = -0.46$, CrI[−0.89, −0.04], $p(\widehat{b} < 0) = 0.983$) and PO sentences ($b = -0.40$, CrI[−0.73, −0.07], $p(\widehat{b} < 0) = 0.992$). As can be seen in Table 3, fixation probabilities were larger for both interest areas in both structures for implausible relative to plausible sentences. This effect was larger for the perceptually relevant region than the control region, such

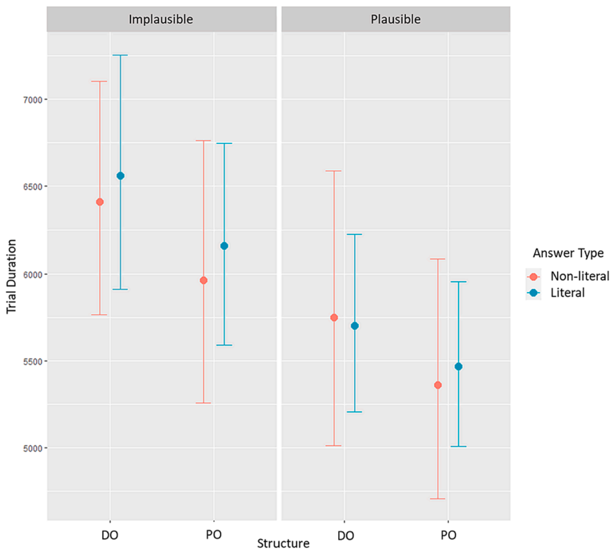

Fig. 2. Effect Estimates Predicting Sentence Reading Time as a Function of Sentence Structure, Plausibility, and whether or not a Noisy-Channel Inference was Made.

that mean fixation probability increased by 0.16 for the perceptually relevant region but only by 0.08 for the control region. There were no reliable effects involving age group in this measure.[4]

Similar patterns were observed for both fixation count and total time. For fixation count, a model using the Poisson family revealed an interaction between plausibility and interest area for both DO ($b = -0.23$, CrI $[-0.37, -0.09]$, $p(\widehat{b} < 0) = 0.999$) and PO sentences ($b = -0.15$, CrI $[-0.29, -0.01]$, $p(\widehat{b} < 0) = 0.982$). Again, the effect of plausibility was markedly elevated for the perceptually relevant region (an increase of 0.49 fixations) compared to the control region (an increase of 0.24 fixations). There was also an interaction between plausibility and age group in PO sentences ($b = -0.17$, CrI$[-0.33, -0.00]$, $p(\widehat{b} < 0) = 0.977$). This was primarily driven by the older adults making fewer fixations in the control (older $m = 0.77$, young $m = 0.88$) and perceptually relevant regions (older $m = 0.81$, younger $m = 1.02$) of plausible sentences, with little difference in the number of fixation they made in the control (older $m = 1.04$, younger $m = 1.08$) and perceptually relevant regions (older $m = 1.39$, younger $m = 1.34$) of implausible sentences.[5]

For total time, there was an interaction between plausibility and interest area for DO sentences ($b = -0.11$, CrI$[-0.20, -0.01]$, $p(\widehat{b} < 0) = 0.984$) with no compelling evidence of an interaction for PO sentences ($b = -0.08$, CrI$[-0.19, 0.03]$, $p(\widehat{b} < 0) = 0.939$). There were effects of plausibility for both DO ($b = -0.16$, CrI$[-0.21, -0.11]$, $p(\widehat{b} < 0) = 1$) and PO sentences ($b = -0.19$, CrI$[-0.25, -0.13]$, $p(\widehat{b} < 0) = 1$). The average time spent in the perceptually relevant region increased by 99 ms for implausible sentences, with this increase being only 56 ms for the control region. Finally, there was an interaction between age group and interest area for DO sentences ($b = -0.13$, CrI$[-0.23, -0.02]$, $p(\widehat{b} < 0) = 0.991$) such that older adults spent longer than young adults fixating the control region (older $m = 389$, young $m = 373$) while young adults spent longer than older adults fixating the perceptually relevant region (older $m = 438$, young $m = 461$).

In summary, we obtained clear effects of sentence plausibility on participants' sampling strategy throughout our sentences, with participants preferentially fixating the perceptually relevant region in response to encountering an implausible sentence. This effect was observed in three measures taking both first-pass and subsequent re-reading behavior into account (i.e., fixation probability, fixation count, total time) but not for a measure taking only first-pass reading into account (i. e., skipping probability). Note that the implausibility of our sentences would not become apparent until the reader reached the end of a sentence, and so it is unsurprising that plausibility effects manifested in eye-movement measures sensitive to re-reading behavior. Age group had limited effects on eye-movement behavior - in only DO sentences did older adults skip more, fixate for longer in the control region, and fixate for less time in the perceptually relevant region than young adults. Finally, in PO sentences young adults made more fixations on the plausible sentences compared to the older adults, with no difference between the two age groups for the implausible sentences.

A final point worth addressing before moving on relates to whether sentence type affected reading behaviour differently depending on whether readers made a noisy-channel inference, like Huang and Staub (2021b) recently observed by for sentences that can be made (un) grammatical by transposing adjacent words. To test this, we constructed an additional model using total sentence reading times as a dependent variable, and examined whether this was affected by sentence plausibility, whether or not participants made a noisy-channel inference, age group, and the interactions between these three factors, nested underneath the effect of sentence structure. If noisy-channel inferences were made rapidly enough that participants never detected the implausibility, then we would expect to observe an interaction between plausibility and whether or not participants made a noisy-channel inference, as outlined above. This model actually revealed main effects of sentence plausibility for both sentence structures (DO: $b = -0.19$, CrI$[-0.23, -0.14]$, $p(\widehat{b} < 0) = 1$; PO: $b = -0.16$, CrI$[-0.22, -0.10]$, $p(\widehat{b} < 0) = 1$), but no interactions with whether or not participants made a noisy-channel inference for either DO ($b = -0.02$, CrI$[-0.12, 0.08]$, $p(\widehat{b} < 0) = 669$) or PO sentences ($b = -0.02$, CrI$[-0.14, 0.09]$, $p(\widehat{b} < 0) = 0.666$). Conditional effects are shown in Fig. 2 – reading times were inflated in implausible sentences regardless of whether readers made a noisy-channel inference. Thus, noisy-channel inferences appear to be revisions to representations derived during initial (first-pass) processing, either during rereading or during reconsideration of previously read text. This model also revealed an interaction between age group and plausibility for the DO sentences only ($b = -0.12$, CrI$[-0.21, -0.03]$, p

[4] It is worth highlighting the fact that the data for our first two measures suggest that our participants would regularly skip our target regions on first-pass reading but then regularly go on to fixate these regions in subsequent re-reading. For example, in our plausible PO sentences the perceptually relevant region only had a probability of being fixated during first-pass reading of 0.27 but had an overall fixation probability of 0.61, meaning that on 34% of the trials on which the region was skipped in first-pass reading it would later be fixated in re-reading. In other words, it was more likely to be fixated in re-reading than during first-pass reading. This could be considered surprising, and potentially indicative of subjects figuring out task demands (i.e., the need to determine whether *to* was absent/present) and so paying special attention to these sentences. To investigate this possibility, we checked whether this pattern remained constant throughout our experiment, finding that this was indeed the case, with, if anything the discrepancy decreasing (i.e., a difference of 0.39 in the first 30 trials and a difference of 0.31 in the last 30 trials). Had the level of re-reading been in response to task demands we would expect the discrepancy to have increased throughout the experiment as task demands became clearer. Furthermore, we found that within our filler items the same pattern of skipping probabilities vs. overall fixation probabilities occurred for regions consisting of three-characters, once again suggesting the pattern was due to standard reading behavior. Thus, while the high level of re-reading might be considered surprising, it does not appear that it occurred due to readers adopting an unnatural reading strategy.

[5] The effects reported for this model were consistent when a range of other priors were used for this model.

Table 4
Mean Literal Interpretation Rates as a Function of Skipping a Region During First-Pass.

	DO		PO	
	PR	Control	PR	Control
Skipped	0.56 (0.02)	0.60 (0.02)	0.80 (0.01)	0.81 (0.01)
Fixated	0.64 (0.03)	0.56 (0.03)	0.82 (0.03)	0.79 (0.03)

Table 5
Mean (Standard Error) Literal Interpretation Rates as a Function of Fixation in a Region (either during First-Pass or Subsequent Re-Reading).

	DO		PO	
	PR	Control	PR	Control
Unfixated	0.55 (0.03)	0.63 (0.03)	0.78 (0.03)	0.80 (0.02)
Fixated	0.60 (0.02)	0.57 (0.02)	0.81 (0.01)	0.81 (0.02)

($\hat{b} < 0) = 0.997$), such that on average there was an 860 ms increase when young adults were reading implausible DO sentences ($m = 6903$ ms) compared to plausible DO sentences ($m = 6043$ ms) while there was an increase of 1430 for older adults (implausible $m = 7121$ ms; plausible $m = 5683$ ms). Thus, it would seem that older adults needed less time to process the plausible sentences than young adults, with both older and young adults taking the same amount of time to process implausible sentences.[6]

The effect of eye-movement behavior on misinterpretations

The above analysis indicates that participants extract more information from areas of uncertainty in text when reading implausible sentences. Next, we focus on whether extracting more information from areas of uncertainty reduces the probability of a reader making a misinterpretation. Such an effect would support the hypothesis that perceptual factors play a role in how likely a reader is to make a noisy-channel inference. To address this question, we constructed Bayesian mixed-effects regression models for the two types of implausible sentences only. As comprehension accuracy for the plausible sentences was near ceiling, there was little point in examining how eye-movement behavior influenced the interpretation of these sentences. Each model reported below treats whether a sentence was interpreted literally or non-literally as a binomial outcome variable. The predictor variables used were sentence structure, with nested simple effects of the eye-movement behavior in each region, the effect of age group, and an interaction between age group and the eye movement behavior in each region. To be clear, reading behavior in the perceptually relevant region and control region were treated as two separate predictors with no interactions with each other.[7] Here, our hypothesis would be confirmed if reading behavior in the perceptually relevant region but not the control region influenced the interpretation that readers derived of the sentences.

There was no compelling evidence for effects of skipping either the perceptually relevant ($b = -0.29$, CrI[-0.70, 0.11], $p(\hat{b} < 0) = 0.924$) or control region ($b = 0.27$, CrI[-0.13, 0.66], $p(\hat{b} > 0) = 0.905$) in DO sentences, with no interaction between age group and skipping in either region (Perceptually Relevant: $b = -0.01$, CrI[-0.74, 0.72], $p(\hat{b} < 0) = 0.506$; Control: $b = 0.16$, CrI[-0.59, 0.90], $p(\hat{b} > 0) = 0.683$). Prepositional object sentences produced no effect of skipping for either the perceptually relevant ($b = -0.14$, CrI[-0.69, 0.39], $p(\hat{b} < 0) = 0.693$) or control region ($b = 0.12$, CrI[-0.45, 0.65], $p(\hat{b} > 0) = 0.671$), with no interactions between age group and skipping behaviour in the perceptually relevant ($b = -0.17$, CrI[-1.10, 0.76], $p(\hat{b} < 0) = 0.645$) or control region ($b = -0.22$, CrI[-1.13, 0.68], $p(\hat{b} < 0) = 0.683$). There was an effect of age group on interpretations for DO sentences ($b = 0.66$, CrI[0.01, 1.31], $p(\hat{b} > 0) = 0.977$), as would be expected on the basis of our earlier analysis focusing on age (See Table 4).

There was little evidence that whether either the perceptually relevant or control region was fixated (during first-pass or subsequent re-reading of the region) affected misinterpretation rates (see Table 5). Specifically, for DO sentences there was no evidence of an effect for the perceptually relevant ($b = 0.14$, CrI[-0.33, 0.60], $p(\hat{b} > 0) = 0.731$) or control region ($b = -0.19$, CrI[-0.65, 0.26], $p(\hat{b} < 0) = 0.799$), with no interaction between age group and fixation probability in the perceptually relevant ($b = -0.02$, CrI[-0.83, 0.78], $p(\hat{b} < 0) = 0.517$) or control region ($b = -0.41$, CrI[-1.23, 0.40], $p(\hat{b} < 0) = 0.843$). Similarly, for PO sentences there was no evidence of an effect for the perceptually relevant region ($b = 0.45$, CrI[-0.17, 1.06], $p(\hat{b} > 0) = 0.927$) or the control region ($b = 0.09$, CrI[-0.51, 0.65], $p(\hat{b} > 0) = 0.628$), with no interaction between age group and fixation probability in the perceptually relevant ($b = -0.16$, CrI[-1.17, 0.87], $p(\hat{b} < 0) = 0.621$) or control region ($b = -0.20$, CrI[-1.11, 0.71], $p(\hat{b} < 0) = 0.673$). There was evidence that older adults made more misinterpretations than the young adults for DO sentences ($b = 0.73$, CrI [0.05, 1.42], $p(\hat{b} > 0) = 0.983$).

There was a clear effect of fixation count on misinterpretation rates for DO sentences (see Fig. 3). Specifically, the more fixations within the perceptually relevant region, the less likely the sentence was misinterpreted on that trial ($b = 0.27$, CrI[0.05, 0.50], $p(\hat{b} > 0) = 0.992$), with no interaction between fixation count and age group ($b = -0.17$, CrI[-0.59, 0.25], $p(\hat{b} < 0) = 0.792$). In contrast, fixation count for the control region had no effect for these sentences ($b = -0.01$, CrI[-0.24, 0.22], $p(\hat{b} < 0) = 0.550$), with no interaction with age ($b = -0.16$, CrI [-0.60, 0.26], $p(\hat{b} < 0) = 0.771$). There was no effect of fixation count in the perceptually relevant region on misinterpretation rates for PO sentences ($b = 0.11$, CrI[-0.24, 0.49], $p(\hat{b} > 0) = 0.716$) with no interaction with age ($b = 0.07$, CrI[-0.54, 0.68], $p(\hat{b} > 0) = 0.582$). There was no evidence of an effect for the control region ($b = 0.25$, CrI [-0.03, 0.52], $p(\hat{b} > 0) = 0.963$) and no interaction with age ($b = -0.03$, CrI[-0.53, 0.48], $p(\hat{b} < 0) = 0.550$). Once again, there was evidence that older adults made more misinterpretations than young

[6] Originally this manuscript included an additional analysis, with the goal of assessing an alternative theory of noisy-channel inferences presented by Cai et al. (2021). In this theory it is assumed that non-literal interpretations of sentences occur due to readers making predictions about syntactic structure on the basis of the thematic role of the first argument of the verb (e.g. in *The mother handed the candle…* readers predict *to* due to candles being inanimate; in *The mother handed the daughter…* readers predict a double-object sentence due to daughters being animate). If this was true, readers would be expected to show faster reading times in sentences where these predictions are confirmed as opposed to contradicted. However, a reviewer pointed out a confound that made it impossible to draw sensible conclusions from this analysis. We mention this analysis as a cautionary tale for future researchers who might look for the same effect that this confound will likely occur in the majority of sentences providing a test of the noisy-channel theory.

[7] For example, in the model for fixation count our model syntax would have been (Interpretation ~ structure/((Fixation count in perceptually relevant region*age group) + (fixation count in control region*age group)) + random effects).

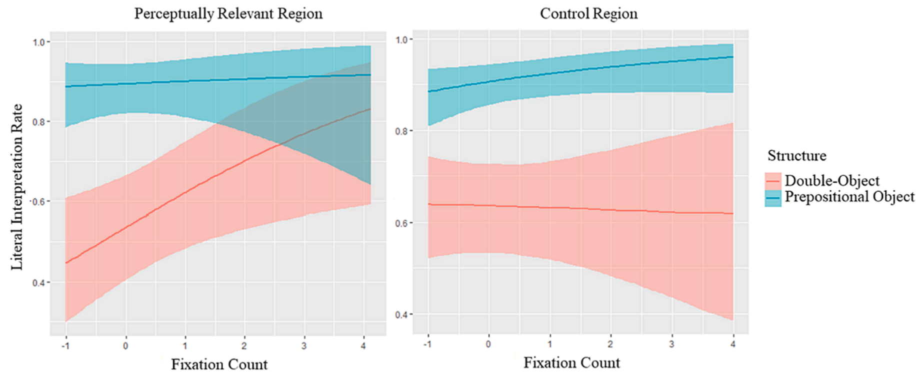

Fig. 3. Effect Estimates Predicting Misinterpretations as a Function of Fixation Count in Perceptually-Relevant (left) and Control Interest Areas (right).

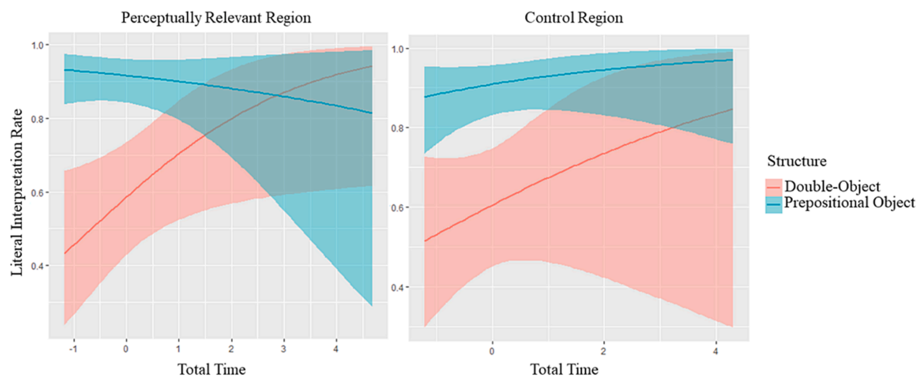

Fig. 4. Effect Estimates Predicting Misinterpretations as a Function of Total Time within Perceptually-Relevant (left) and Control Interest Areas (right).

adults for DO sentences ($b = 0.78$, CrI[0.12, 1.44], $p(\widehat{b} > 0) = 0.990$).

When total time within a region was the predictor variable, the model revealed that the longer participants spent in the perceptually relevant region of DO sentences, the less likely they were to misinterpret the sentence ($b = 0.37$, CrI[0.01, 0.77], $p(\widehat{b} > 0) = 0.978$; see Fig. 4), with no corresponding effect for the control region ($b = 0.14$, CrI[−0.26, 0.57], $p(\widehat{b} > 0) = 0.741$). For DO sentences, age group interacted with neither total time in the perceptually relevant ($b = -0.31$, CrI[−1.00, 0.39], $p(\widehat{b} < 0) = 0.815$) nor control region ($b = -0.34$, CrI[−1.19, 0.47], $p(\widehat{b} < 0) = 0.788$). For the PO sentences total time within neither the control region ($b = 0.40$, CrI[−0.03, 0.89], $p(\widehat{b} > 0) = 0.965$) nor perceptually relevant region ($b = 0.03$, CrI[−0.39, 0.49], $p(\widehat{b} > 0) = 0.537$) had an effect on sentence interpretation, with again no interaction between age group and total time in the perceptually relevant ($b = 0.43$, CrI[−0.32, 1.20], $p(\widehat{b} > 0) = 0.873$) or control region ($b = 0.25$, CrI[−0.53, 1.06], $p(\widehat{b} > 0) = 0.729$). Finally, it is worth noting that for this model there was no effect of age group on interpretation of DO sentences ($b = 0.61$, CrI[−0.17, 1.39], $p(\widehat{b} > 0) = 0.939$). This might seem odd given the presence of this effect in the models in which age group was included alongside our other three measures. However, recall that for our total time data we excluded instances with a total time of 0 ms. As such, the power of this model to detect the age group effect on interpretations was reduced.

While we have primarily adopted the approach of examining reading behaviour in our pairs of perceptually relevant and control regions, we also analysed our data using an approach similar to Staub et al. (2019). Here, instead of focusing on reading behaviour in our somewhat unconventional interest areas, this analysis examines reading behaviour on the word preceding and following the site of the potential edit within the sentence, such that we consider whether people were less likely to make a noisy-channel inference when they directly fixated both of these words in first-pass reading as opposed to when they only fixated one or neither of these words. To test this possibility, we constructed a further Bayesian mixed-effects model which treated whether or not readers made a noisy-channel inference for implausible sentences as a dependent variable, and tested for an effect of whether participants fixated both the preceding and following word or just one or neither of these two words, nested underneath sentence structure. The pattern of fixations was coded to interact with age group in the model. While there was a trend in our data towards an effect of fixation pattern for both the DO (probability of a noisy-channel inference after fixating both words = 0.36; after fixating one or neither of these words = 0.45) and PO sentences (0.17 after fixating both; 0.23 after fixating one or neither) these effects were not reliable in our statistical model (DO: $b = -0.31$, CrI[−0.71, 0.08], $p(\widehat{b} < 0) = 0.941$; PO: $b = -0.36$, CrI[−0.88, 0.18], $p(\widehat{b} < 0) = 0.909$). Once again, there was a main effect of age on the interpretation of DO sentences ($b = -0.63$, CrI[−1.26, −0.00], $p(\widehat{b} < 0) = 0.975$).

Discussion

Our aim with the present study was to link eye-movement behavior and the extraction of visual information during reading to predictions derived from noisy-channel inference accounts of why people systematically misinterpret certain implausible sentences (Gibson et al., 2013; 2017; Ryskin et al., 2018). Specifically, we were interested in 1) whether readers selectively extract more information from a site at which a change in perceptual input would make implausible sentences plausible (Cutter et al., 2022b; Levy et al., 2009); 2) whether eye-movement behavior at these locations affects the probability of a reader making a noisy-channel inference; and 3) whether changes in visual and cognitive function inherent in healthy aging (Hartshorne & Germine, 2015; Kaufman et al., 2016; Owsley, 2011; Salthouse, 2010) increase the

Table A1
Output for the model looking at the effect of age on the probability of readers making noisy-channel inferences, taking account of the animacy of the referent in the comprehension question. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	b	L95	U95	$p(\widehat{b} >/< 0)$
Intercept	**2.72**	**2.35**	**3.12**	**1**
Plausibility	**2.42**	**1.83**	**3.02**	**1**
Structure	**1.33**	**0.89**	**1.79**	**1**
Age	−0.20	−0.81	0.42	0.736
Plausibility*Structure	**−0.80**	**−1.48**	**−0.13**	**0.990**
Plausibility*Age	**0.97**	**0.15**	**1.79**	**0.990**
Structure*Age	0.39	−0.25	1.03	0.882
Plausibility*Structure*Age	0.41	−0.59	1.39	0.791
Animacy	0.20	−0.23	0.63	0.820
Animacy*Plausibility	**1.89**	**0.87**	**2.88**	**0.999**
Animacy*Structure	0.41	−0.24	1.08	0.892
Animacy*Age	−0.05	−0.65	0.56	0.562
Animacy*Plausibility*Structure	0.80	−0.28	1.87	0.930
Animacy*Plausibility*Age	0.18	−1.12	1.49	0.610
Animacy*Structure*Age	0.35	−0.57	1.26	0.767
Animacy*Plausibility*Age*Structure	0.67	−0.80	2.15	0.814

Table A2
Output for the model looking at how long readers spend reading sentences as a function of plausibility and eventual interpretation. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	B	L95	U95	$p(\widehat{b} >/< 0)$
Intercept	**8.65**	**8.59**	**8.72**	**1**
Structure	−0.03	−0.07	0.01	0.943
Double-Object				
Plausibility	**−0.19**	**−0.23**	**−0.14**	**1**
Interpretation	**0.05**	**−0.01**	**0.10**	**0.959**
Age	−0.00	−0.12	0.11	0.513
Plausibility*Interpretation	−0.02	−0.12	0.08	0.669
Plausibility*Age	**−0.12**	**−0.21**	**−0.03**	**0.997**
Interpretation*Age	0.08	−0.01	0.17	0.948
Plausibility*Interpretation*Age	0.02	−0.18	0.22	0.572
Prepositional-Object				
Plausibility	**−0.16**	**−0.22**	**−0.10**	**1**
Interpretation	−0.02	−0.08	0.04	0.758
Age	0.06	−0.05	0.17	0.861
Plausibility*Interpretation	−0.02	−0.14	0. 09	0.666
Plausibility*Age	−0.10	−0.22	0.01	0.962
Interpretation*Age	−0.09	−0.19	0.01	0.956
Plausibility*Interpretation*Age	−0.02	−0.25	0.19	0.578

likelihood of making noisy-channel inferences and how this interacts with eye-movement behaviour. We focus on the initial two questions first, before discussing the effect of age.

Eye-movement behavior and noisy-channel inferences

Our findings were consistent with several predictions we made for noisy-channel processing accounts. To start with the question of whether readers selectively inspect a location at which a word could be inserted or removed to make an implausible sentence more plausible, our data offers clear evidence that this is indeed the case. Participants spent longer reading implausible than plausible sentences. Furthermore, more time was spent reading the site of a potential edit than another spatially equivalent but irrelevant control region. Specifically, participants were more likely to fixate this region, both making more fixations and fixating for longer at this region of implausible sentences. Accordingly, participants clearly were focusing on sites of potential edits when faced with implausible material. No such effect was observed in region skipping behavior. However, this is unsurprising as, when a reader first encounters the critical region, they would be unaware whether the sentence was implausible. The lack of a skipping effect therefore reinforces the notion that effects observed in other measures were due to sentence implausibility, as opposed to artefacts relating to material preceding the implausibility.

Table A3
Output for the model looking at the effect of sentence structure, plausibility, interest area, and age on skipping probability. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	b	L95	U95	$p(\widehat{b} >/< 0)$
Intercept	**0.90**	**0.75**	**1.05**	**1**
Structure	**0.44**	**0.28**	**0.59**	**1**
Double-Object				
Plausibility	0.00	−0.16	0.16	0.485
Interest Area	−0.07	−0.35	0.21	0.692
Age	**0.29**	**0.03**	**0.55**	**0.986**
Plausibility*Interest Area	−0.01	−0.29	0.27	0.531
Plausibility*Age	−0.03	−0.38	0.32	0.571
Interest Area*Age	0.20	−0.19	0.59	0.850
Plausibility*IA*Age	0.58	−0.03	1.19	0.969
Prepositional-Object				
Plausibility	0.03	−0.14	0.20	0.649
Interest Area	0.05	−0.21	0.30	0.645
Age	0.13	−0.14	0.41	0.835
Plausibility*Interest Area	0.10	−0.24	0.42	0.724
Plausibility*Age	0.08	−0.23	0.40	0.697
Interest Area*Age	−0.13	−0.59	0.32	0.720
Plausibility*IA*Age	0.35	−0.26	0.97	0.872

Table A4
Output for the model looking at the effect of sentence structure, plausibility, interest area, and age on fixation probability. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	b	L95	U95	$p(\widehat{b} >/< 0)$
Intercept	**0.84**	**0.66**	**1.02**	**1**
Structure	**−0.20**	**−0.39**	**−0.00**	**0.977**
Double-Object				
Plausibility	**−0.60**	**−0.82**	**−0.39**	**1**
Interest Area	−0.03	−0.30	0.24	0.586
Age	0.04	−0.28	0.36	0.597
Plausibility*Interest Area	**−0.46**	**−0.89**	**−0.04**	**0.983**
Plausibility*Age	0.12	−0.31	0.56	0.712
Interest Area*Age	−0.20	−0.58	0.18	0.847
Plausibility*IA*Age	−0.05	−0.71	0.62	0.558
Prepositional-Object				
Plausibility	**−0.66**	**−0.85**	**−0.47**	**1**
Interest Area	**0.49**	**0.24**	**0.74**	**0.999**
Age	−0.11	−0.43	0.20	0.760
Plausibility*Interest Area	**−0.40**	**−0.73**	**−0.07**	**0.992**
Plausibility*Age	−0.32	−0.69	0.07	0.949
Interest Area*Age	−0.23	−0.67	0.21	0.854
Plausibility*IA*Age	−0.41	−1.00	0.17	0.918

We now consider the relationship between eye-movement behavior and misinterpretation rates. For implausible double-object sentences – in which the insertion of *to* would make the sentence plausible – participants were more likely to form a literal interpretation when they made more fixations on, and fixated for longer, the region at which *to* could have been inserted, with negligible effects at the control region. This is clear evidence in line with the noisy-channel account, and specifically with the idea advanced in the current paper that greater perceptual certainty would reduce the probability of a misinterpretation. This contradicts the theoretical position advanced in recent research proposing that perceptual factors play no role in noisy-channel

Table A5
Output for the model looking at the effect of sentence structure, plausibility, and interest area on fixation count. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	b	L95	U95	$p(\widehat{b} >/< 0)$
Intercept	−0.00	−0.09	0.08	0.535
Structure	**−0.14**	**−0.21**	**−0.06**	**0.999**
Double-Object				
Plausibility	**−0.31**	**−0.40**	**−0.22**	**1**
Interest Area	0.06	−0.06	0.18	0.845
Age	−0.02	−0.16	0.13	0.585
Plausibility*Interest Area	**−0.23**	**−0.37**	**−0.09**	**0.999**
Plausibility*Age	−0.04	−0.24	0.17	0.641
Interest Area*Age	−0.11	−0.26	0.04	0.929
Plausibility*IA*Age	−0.17	−0.42	0.08	0.912
Prepositional-Object				
Plausibility	**−0.34**	**−0.44**	**−0.24**	**1**
Interest Area	**0.18**	**0.08**	**0.28**	**0.999**
Age	−0.11	−0.26	0.05	0.912
Plausibility*Interest Area	**−0.15**	**−0.29**	**−0.01**	**0.982**
Plausibility*Age	**−0.17**	**−0.33**	**−0.00**	**0.977**
Interest Area*Age	−0.03	−0.21	0.15	0.618
Plausibility*IA*Age	−0.15	−0.42	0.11	0.874

Table A6
Output for the model looking at the effect of sentence structure, plausibility, and interest area on total time. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	b	L95	U95	$p(\widehat{b} >/< 0)$
Intercept	5.77	5.72	5.81	1
Structure	**−0.08**	**−0.13**	**−0.03**	**0.999**
Double-Object				
Plausibility	−0.16	−0.21	−0.11	**1**
Interest Area	**−0.19**	**−0.25**	**−0.13**	**0.999**
Age	−0.02	−0.10	0.07	0.652
Plausibility*Interest Area	**−0.11**	**−0.20**	**−0.01**	**0.984**
Plausibility*Age	−0.10	−0.21	0.01	0.959
Interest Area*Age	**−0.13**	**−0.23**	**−0.02**	**0.991**
Plausibility*IA*Age	−0.15	−0.33	0.04	0.937
Prepositional-Object				
Plausibility	**−0.19**	**−0.25**	**−0.13**	**1**
Interest Area	**0.16**	**0.09**	**0.22**	**1**
Age	−0.06	−0.15	0.03	0.910
Plausibility*Interest Area	−0.08	−0.17	0.02	0.939
Plausibility*Age	−0.08	−0.19	0.03	0.918
Interest Area*Age	−0.02	−0.15	0.11	0.634
Plausibility*IA*Age	0.05	−0.15	0.25	0.689

inference making (Liu et al., 2021). By comparison, this aspect of our findings suggests there is a relationship between eye movement behavior during reading and the likelihood of misinterpreting implausible sentences, certainly for the double-object sentences that we examined.

Such effects were not observed for the implausible prepositional object sentences, however. For these sentences, reading behavior in neither the control nor the perceptually relevant region affected misinterpretation rates. These findings for the prepositional object sentences might appear to be paradoxical. Clearly, readers spent longer fixating within relevant regions when faced with the implausible form of these sentences as opposed to the plausible form. However, the extra perceptual information that was presumably extracted in this time did nothing to decrease the probability of participants misinterpreting these sentences when we examined participants' interpretations as a function of eye-movement behaviour at the perceptually relevant region. It is

Table A7
Output for the model looking at the effect of interest area skipping on implausible sentence interpretation. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	b	L95	U95	$p(\widehat{b} >/< 0)$
Intercept	**1.39**	**0.98**	**1.81**	**1**
Structure	**1.66**	**1.23**	**2.13**	**1**
Double-Object				
Skipped Perceptually Relevant	−0.29	−0.70	0.11	0.924
Skipped Control	0.27	−0.13	0.66	0.905
Age	**0.66**	**0.01**	**1.31**	**0.977**
Skipped Perceptually Relevant*Age	−0.01	−0.74	0.72	0.506
Skipped Control*Age	0.16	−0.59	0.90	0.665
Prepositional-Object				
Skipped Perceptually Relevant	−0.14	−0.69	0.39	0.693
Skipped Control	0.12	−0.45	0.65	0.671
Age	0.59	−0.21	1.38	0.683
Skipped Perceptually Relevant*Age	−0.17	−1.10	0.76	0.645
Skipped Control*Age	−0.22	−1.13	0.68	0.929

Table A8
Output for the model looking at the effect of whether or not each interest area was fixated during either first-pass or subsequent re-reading on implausible sentence interpretation. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	B	L95	U95	$p(\widehat{b} >/< 0)$
Intercept	**1.36**	**0.94**	**1.79**	**1**
Structure	**1.59**	**1.14**	**2.08**	**1**
Double-Object				
Fixated Perceptually Relevant	0.14	−0.33	0.60	0.731
Fixated Control	−0.19	−0.65	0.26	0.799
Age	**0.73**	**0.05**	**1.42**	**0.983**
Fixated Perceptually Relevant*Age	−0.02	−0.83	0.78	0.517
Fixated Control*Age	−0.41	−1.23	0.40	0.843
Prepositional-Object				
Fixated Perceptually Relevant	0.45	−0.17	1.06	0.927
Fixated Control	0.09	−0.51	0.65	0.628
Age	0.61	−0.19	1.43	0.934
Fixated Perceptually Relevant*Age	−0.16	−1.17	0.87	0.621
Fixated Control*Age	−0.20	−1.11	0.71	0.673

worth considering, however, the relative likelihood that incoming information has a missing word (as in our DO sentences) versus an erroneously inserted word (as in our PO sentences). If we assume that the former is more likely, then it is arguably more rational to base one's interpretation on evidence of a possible deletion than a possible insertion (as there is more uncertainty in the potential deletion condition than the potential insertion condition). Furthermore, readers are also likely to be more certain about the presence of *to* in PO sentences having perceived it at any point during reading, while having not perceived *to* in an initial pass of the DO sentences would not guarantee its absence. As such, the additional evidence acquired during re-reading would be expected to be more useful for DO than PO sentences. These distinctions are already made within Gibson et al.'s theoretical approach, via the Bayesian size principle (MacKay, 2003). Thus, the absence of an effect of how long readers spent at the site of the potential edit for PO sentences contradicts neither the noisy-channel framework in general, nor the more specific claim made in the current paper that eye-movements play a role in how readers form a noisy-channel inference.

It is interesting to note that, even when participants acquired an overwhelming level of perceptual evidence for the double-object sentences, there was still a relatively high probability of them

Table A9
Output for the model looking at the effect of how many fixations each interest area received on sentence interpretation. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	B	L95	U95	$p(\hat{b} >/< 0)$
Intercept	**1.44**	**1.01**	**1.87**	**1**
Structure	**1.82**	**1.38**	**2.29**	**1**
Double-Object				
Fixation Count Perceptually Relevant	**0.27**	**0.05**	**0.50**	**0.992**
Fixation Count Control	−0.01	−0.24	0.22	0.550
Age	**0.78**	**0.12**	**1.44**	**0.990**
Fixation Count Perceptually Relevant*Age	−0.17	−0.59	0.25	0.792
Fixation Count Control*Age	−0.16	−0.60	0.26	0.771
Prepositional-Object				
Fixation Count Perceptually Relevant	0.11	−0.24	0.49	0.716
Fixation Count Control	0.25	−0.03	0.52	0.963
Age	0.49	−0.29	1.28	0.893
Fixation Count Perceptually Relevant*Age	0.07	−0.54	0.68	0.582
Fixation Count Control*Age	−0.03	−0.53	0.48	0.550

Table A10
Output for the model looking at the effect of how long each interest area was fixated for on sentence interpretation. Effects for which the 95 % credible interval does not contain 0 are presented in bold.

Effect	b	L95	U95	$p(\hat{b} >/< 0)$
Intercept	**1.58**	**1.10**	**2.10**	**1**
Structure	**1.94**	**1.40**	**2.53**	**1**
Double-Object				
Total Time Perceptually Relevant	**0.37**	**0.01**	**0.77**	**0.978**
Total Time Control	0.14	−0.26	0.57	0.741
Age	0.61	−0.17	1.39	0.939
Total Time Perceptually Relevant*Age	−0.31	−1.00	0.39	0.815
Total Time Control*Age	−0.34	−1.19	0.47	0.788
Prepositional-Object				
Total Time Perceptually Relevant	0.03	−0.39	0.49	0.537
Total Time Control	0.40	−0.03	0.89	0.965
Age	0.38	−0.52	1.29	0.796
Total Time Perceptually Relevant*Age	0.43	−0.32	1.20	0.873
Total Time Control*Age	0.25	−0.53	1.06	0.729

Table A11
Output for the model looking at the effect of whether readers fixated both words surrounding the potential edit vs. just one or neither words on sentence interpretation.

Effect	b	L95	U95	$p(\hat{b} >/< 0)$
Intercept	1.37	0.97	1.78	1
Structure	1.59	1.19	2.01	1
Double-object				
Fixation Pattern	−0.31	−0.71	0.08	0.941
Age	**−0.63**	**−1.26**	**−0.00**	**0.976**
Fixation Pattern*Age	−0.13	−0.85	0.60	0.633
Prepositional object				
Fixation Pattern	−0.36	−0.88	0.18	0.909
Age	−0.44	−1.19	0.32	0.875
Fixation Pattern*Age	0.46	−0.42	1.35	0.852

misinterpreting these sentences (i.e., three fixations on the perceptually relevant region were still accompanied by a 36 % misinterpretation rate). This paradox can be explained by the fact that, within the noisy-channel account, it is assumed that a large proportion of inferences are made as a result of comprehenders assuming that a mistake was made in the production of the sentence, rather than in their own perception. According to this account, comprehenders make misinterpretations due to assuming that something was miswritten or misspoken, with the writer having a different intended meaning (e.g. Ryskin et al., 2018). For some implausible sentences, participants might use extra perceptual information to confirm that the sentence is anomalous, then assume that this was an error in production and so interpret it according to what they believe was intended. Regardless of the veracity of this explanation, it is clear that perceptual factors play an important role in how likely a reader is to misinterpret that sentence, contrary to the proposal put forward in Liu et al. (2021).

It should also be noted that misinterpretation rates for our PO sentences were low compared to those in prior studies. Specifically, for implausible PO sentences, we observed a misinterpretation rate of only 19 %, whereas Gibson et al. (2013) observed a ∼ 40 % misinterpretation rate. This is interesting for several reasons. Hypothetically, it could be that, for the PO sentences, a particular pattern of eye-movements would result in readers making more misinterpretations for implausible sentences, but we did not observe this pattern. Accordingly, before concluding that misinterpretation rates for these sentences are not influenced by eye-movement behavior, it will be valuable to confirm the current findings in a participant sample that exhibits a higher baseline error rate. It is important to note that there are differences in the paradigm used here and that used by Gibson et al. (2013) which may have led to differences in how likely participants were to make a noisy-channel inference. In our study, participants only saw the comprehension probe after finishing reading a sentence, whereas in Gibson et al. (2013) the probe was displayed on screen at the same time as the sentence. One might expect that the simultaneous presence of the comprehension probe and sentence would allow for extra confirmatory re-reading of the sentence and thus higher correct performance (as reflected in higher levels of literal interpretations). On the other hand, the format of the comprehension questions was also somewhat 'unusual'. Accordingly, if participants are simultaneously presented with a sentence and a question which are both implausible, this might instead lead to an increase in confusion, potentially resulting in more non-literal interpretations. Future work may benefit from testing a single group of participants on both questioning styles to determine if this effect is simply an artifact of different participants, or if processing fundamentally differs. The latter possibility may be problematic for noisy-channel inference accounts. Finally, differences in the experimental procedures across these studies may have contributed to the differences in participants' behavior. Specifically, it may be the case that the more 'formal' nature of the eye-tracking study – in which the experimenter is in the room with the participant and observing their behavior as the experiment progresses etc., compared with previous internet-based survey data collection techniques, may lead participants to expect it to be more unlikely that there are errors in the stimuli in the former case. Consequently, readers may make fewer misinterpretations due to believing that the literal interpretation of the sentence is more likely to truly represent the intended meaning.

We also examined whether readers showed signs of reading disruption on trials in which they made a noisy-channel inference. Our analysis showed that the level of reading disruption was not reliably different on trials in which readers made a noisy-channel inference relative to trials on which they did not. This contrasts with the Huang and Staub (2021b) findings, in which no reading difficulty was observed for sentences in which two words were transposed (e.g. *the white was dog big*) when participants failed to report that these sentences were ungrammatical. Huang and Staub attributed this effect to readers making a noisy-channel correction to the order of *was* and *dog* prior to integrating these words into the unfolding syntactic representation. In our stimuli, the necessary edits to the input (i.e., the deletion/insertion of *to*) would not have become apparent to readers until they understood the sentence as a whole, meaning that any edits could only occur once the implausibility had been detected. Thus, in our study, it would make sense that disruption was observed even when participants interpreted the sentences as a plausible alternative. By comparison, no disruption may have been observed in Huang and Staub's study when a noisy-channel inference was made, because this would have involved the reader making an edit to the input before this was integrated with the rest of the syntactic structure.

Our finding of targeted re-reading behavior for implausible sentences may support proposals by Levy (2008; Levy et al., 2009). Levy proposed that readers maintain uncertainty about prior perceptual input, revising the current sentence representation to a previously less probable structure when presented with information that renders the previously more probable parse unlikely. The principal existing evidence for this proposal (Levy et al., 2009) has recently been shown to not replicate (Cutter et al., 2022a, 2022b), with additional independent evidence (Paape et al., 2022) questioning this account. If the current findings were interpreted as supporting Levy's proposal, this would represent novel evidence for a theory that contradicts dominant approaches to syntactic parsing. These traditional accounts assume readers rapidly commit to a single syntactic analysis of an ambiguity rather than maintaining uncertainty and continually updating their assessment of which possible parse is most likely in a probabilistic manner as new information becomes available (e.g., van Gompel et al., 2001). According to Levy's proposal, the re-reading aspect of our findings would be due to readers considering both a DO and PO interpretation simultaneously, and, on encountering the implausibility, updating the probabilities towards the structure for which they have less evidence due to the presence/absence of *to*. Re-reading behavior would be due to readers checking whether this word was truly present/absent to adjudicate between the two parses which are both unlikely for different reasons. A note of caution is necessary here, however. It could be that, rather than maintaining uncertainty as they read, readers might begin considering alternative parses only once they encounter an implausibility, and then seek out relevant perceptual information. Further work is needed to adjudicate between these two possibilities. Moreover, future work should investigate whether evidence for word-level uncertainty, as originally shown by Levy et al. (2009), can be found for sentences that are implausible as opposed to the merely syntactically unlikely structures examined in prior studies (see also Cutter et al., 2022b).

The present findings are also relevant to the more general question of how the oculomotor control system interacts with processes guiding the reinterpretation of sentences. Some studies (Frazier & Rayner, 1982; Meseguer et al., 2002; Mitchell et al., 2008; von der Malsburg & Vasishth, 2011) have examined how readers attempt to syntactically reanalyze garden-path sentences (e.g., *The babysitter bought a gift card thanked the parents*) upon realising they made an attachment error earlier in the sentence. This work suggests that readers follow a systematic reanalysis strategy, by targeting regressive eye-movements back to the site of the attachment error (i.e., *bought a gift card*), as opposed to simply back to the start of the sentence to allow readers to re-analyse the entire structure. Our own work extends this finding by showing that reading behavior is also targeted selectively under conditions in which the sentence meaning is implausible, causing readers to seek out potential perception errors. Christianson et al., (2017; see also Schotter et al., 2014) attempted to link re-reading behavior in garden-path sentences to the probability of readers making a 'good-enough' interpretation of these sentences; that is, assigning a plausible but incorrect meaning to a syntactic structure, whereby readers interpret that the babysitter was the one who purchased the gift card (Christianson et al., 2001; Christianson et al., 2006; Ferreira et al., 2001). They observed no effect of re-reading on the eventual interpretations of these sentences. At a surface level this finding may be considered inconsistent with our work, in which reading behavior did affect misinterpretation rates in double-object sentence structures. However, Christianson et al. (2017) argued that their null effect was due to re-reading being used to confirm as opposed to revise participants' initial (mis)interpretation of the garden-path sentences. This is consistent with our findings, as confirmatory re-reading within our sentences leads to reinforcement of the correct initial literal interpretation, hence decreasing misinterpretation rates. This effect could occur either in instances when participants initially generate a literal interpretation (and thus reinforce this) or generate a non-literal interpretation and then revert back to a literal interpretation upon confirming there is no evidence for the edit required for the non-literal interpretation.

A further consideration might be how aware participants were of reading implausible sentences, as well as the oddities involved in the comprehension questions (e.g., *Did the candle receive someone/something?*). When debriefing our participants, we queried whether they had noticed anything odd about the experimental materials. All participants who were asked this question indicated that they had detected oddities, with these being specific to the experimental sentences rather than the filler items.[8] Participants varied in terms of their response, with some reporting that inanimate objects were receiving people, which seemed odd. Others were more specific in stating that the word order felt wrong or that sentences would make more sense with the word *to* inserted or removed. An interesting question arising from this consideration concerns whether reading behavior differs between participants who are aware of the relevance of the presence/absence of *to* versus those who are not. The former group might be more likely to adopt an eye movement strategy that involves targeting this region to test for perception errors. This could be explored in future work with a larger sample of participants and more systematic debriefing process. Future work might also employ a more nuanced measure of interpretation, such as including confidence ratings (see Christianson et al., 2001 for a similar approach to garden-path misinterpretations), to assess participants' awareness of how they are processing the implausible sentences.

There is one final issue we wish to address regarding the link between eye-movement behaviour and noisy-channel inference making, and something which could be viewed as a flaw in our analytical approach. It could be argued that the misinterpretation effects observed in our study — and in prior investigations of noisy-channel inferences — are not exclusively due to misinterpretation of the target sentence, but rather that there is also a possibility of readers misinterpreting the questions within the noisy-channel framework. The argument here would be that participants sometimes mentally edit the question such that instead of reading it as "Did the candle receive something/someone?" they read it as "Did something/someone receive the candle?". Such word reversals are considered to be noisy-channel edits. If readers were indeed making these edits, it would be problematic for our approach, since in this case we should have also examined eye-movement behavior on the comprehension questions as well as our target sentences. However, on the basis of prior literature, we find the possibility that noisy-channel edits of the comprehension questions are driving our misinterpretation effects to be exceedingly unlikely, for several reasons. First, in Gibson et al.'s (2013) original study a range of sentence constructions were examined. This included the DO/PO sentences used in the current study as the sentence type which would require the smallest edit (i.e., the insertion/deletion of *to*) and at the other extreme active/passive sentences (e.g., *The ball kicked the girl; The girl was kicked by the ball*) which would require the largest edit (i.e., the insertion/deletion of both *was* and *by*). Both sentence types were

[8] Not all participants were asked these questions, with it only becoming apparent that it would be worthwhile recording the subjective experience of reading the sentences after it became clear how aware some participants were of the unusual sentences. Given that not all participants were debriefed in the same manner we avoid presenting any formal analysis of this data.

followed by the same format of question, and thus if misinterpretation effects were driven by noisy-channel edits to the comprehension question, Gibson et al. should have observed a similar misinterpretation rate for both sentence types. This is not what was found; rather, in Experiment 1 misinterpretation for the DO/PO sentences occurred on between 40 % and 50 % of trials, whereas for active/passive sentences they occurred on less than 5 % of trials. This suggests that misinterpretations are driven predominantly if not entirely by the target sentences as opposed to the comprehension questions that follow them. Furthermore, subsequent studies have found noisy-channel misinterpretation effects when interpretation is assessed via a picture-matching task (Warren et al., 2017), a task in which participants act out their interpretation of the sentence using a set of dolls (Gibson et al., 2015), or even syntactic priming (Cai et al., 2022) as opposed to a comprehension question. Indeed, in the syntactic priming study by Cai et al. equivalent priming effects were found regardless of whether the extent of syntactic priming was assessed between when participants heard the implausible sentence and answered a comprehension question, or after participants heard the implausible sentence and then answered the comprehension question. If misinterpretation of the comprehension questions were contributing to the current phenomenon in any meaningful manner, then one would expect to observe a modulation of the syntactic priming effect by whether comprehension was probed before or after syntactic priming. For all of these reasons, we consider it unlikely that readers were making misinterpretations due to making noisy-channel edits to the comprehension questions.

Aging and noisy-channel inferences

We also considered whether cognitive aging leads readers to make more noisy-channel inferences. We hypothesized that declines in visual processing (Owsley, 2011) and increases in linguistic knowledge (Hartshorne & Germine, 2015) in older adults might lead them to make more noisy-channel inferences than young adults. The older adults in the present study exhibited the expected differences in visual processing and linguistic ability. We therefore should have observed their effects if they genuinely influence noisy-channel processing. Accordingly, we did observe evidence of an age group difference, such that misinterpretation rates were more strongly influenced by plausibility for the older than the younger adults. Therefore, we do have evidence that older adults are more likely to make noisy-channel inferences. It would be of interest to conduct future work using a larger sample to allow for a breakdown of how specific individual differences (i.e., lower acuity, increased linguistic knowledge) influence noisy-channel inference making.

Age group had limited effects on eye-movement behavior, such that in only DO sentences did older adults skip more, fixate for longer in the control region, and fixate for less time in the perceptually relevant region than young adults. In PO sentences, young adults made more fixations in the plausible sentences than older adults, with no real difference between the two age groups for the implausible sentences. In sentence reading times, older adults spent less time reading plausible DO sentences than young adults, with little difference between the age groups for the implausible sentences. The only one of these effects which might be relevant to noisy-channel inference making is older adults' propensity to spend less time in the perceptually relevant region of DO sentences than young adults. Here, it could be that this in part accounts for the older adults making more noisy-channel inferences than the young adults, due to having less perceptual evidence concerning where *to* could be inserted. None of the other effects seem likely to have driven differences in interpretation between the two groups. Furthermore, in terms of how eye-movement behavior affected interpretation in each age group there was no compelling evidence for interactions between age group and eye-movement behaviour in any of our models, with only main effects of age group on interpretations of DO sentences. The fact that there was an effect of age group on interpretation even in models which took account of eye movement behavior suggests that these age group effects could not purely have been driven by the two groups extracting different levels of perceptual information from the sentences. Furthermore, the lack of interactions in these models suggest that any greater uncertainty older adults place upon the perceptual information they extract during reading is based upon the totality of that information, as opposed to each separate piece.

It is also worth noting that participants were more likely to make noisy-channel inferences for sentences in which the referent in the comprehension question was animate as opposed to inanimate. This effect was most likely because, in the absence of any other information, it seems more believable that an animate referent would receive something than an inanimate entity. Future work could follow up this possibility by more systematically examining the effect of question plausibility in noisy-channel inference making.

Conclusion

We set out to obtain novel evidence regarding how the oculomotor control system interfaces with sentence comprehension mechanisms, by examining eye-movement behavior for sentences considered to trigger noisy-channel inferences. Our findings suggest that, when presented with implausible linguistic input, the oculomotor control system is used to seek out evidence that might allow the reader to arrive at a plausible interpretation or to confirm the literal implausible interpretation. Specifically, we find that when inserting a word would support a more plausible interpretation, the more time that readers dwell on this potential insertion location, the less likely they are to conclude that their percept of the sentence is faulty. This finding represents important evidence that noisy-channel inferences for implausible sentences are not only made due to comprehenders assuming an error has been made in the production of a sentence, but also due to their own perception of that sentence. We additionally observed evidence suggesting that age-related changes in visual and cognitive systems may increase the propensity for readers to make noisy-channel inferences. Crucially, these findings have implications for the workings of the oculomotor control system, and for the wider psycholinguistic literature regarding the highly influential noisy-channel approach to language processing.

CRediT authorship contribution statement

Michael G. Cutter: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Kevin B. Paterson:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Ruth Filik:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by the Leverhulme Trust Research Project Grant RPG-2019-051.

These findings were previously presented at the 2022 CUNY Conference on Human Sentence Processing and to the Experimental Psychology Society July 2022 meeting.

Appendix A

This appendix contains tables which present the output of each of our Bayesian mixed-effect regression models. It should be noted that for the directional probability we provide the probability of the effect being in the more likely direction. Tables A1–A11.

References

Balota, D. A., & Rayner, K. (1991). Word recognition processes in foveal and parafoveal vision: The range of influences of lexical variables. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 198–232). Hillsdale, NJ: Erlbaum.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53-79). https://core.ac.uk/download/pdf/55852074.pdf.

Bürkner, P. C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal, 10*, 395–411. https://doi.org/10.32614/RJ-2018-017

Cai, Z. G., Zhao, N., & Pickering, M. J. (2021). How do people interpret implausible sentences? *PsyArxiv*. https://doi.org/10.31234/osf.io/cnzxq

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

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes, 42*, 205–238. https://doi.org/10.1207/s15326950dp4202_6

Christianson, K., Luke, S. G., Hussey, E. K., & Wochna, K. L. (2017). Why reread? Evidence from garden-path and local coherence structures. *Quarterly Journal of Experimental Psychology, 70*, 1380–1405. https://doi.org/10.1080/17470218.2016.1186200

Cutter, M. G., Filik, R., & Paterson, K. B. (2022). Do readers maintain word-level uncertainty during reading? A pre-registered replication study. *Journal of Memory and Language, 125*, Article 104336. https://doi.org/10.1016/j.jml.2022.104336

Cutter, M. G., Paterson, K. B., & Filik, R. (2022). No evidence of word-level uncertainty in younger and older adults in self-paced reading. *Quarterly Journal of Experimental Psychology, 75*, 1085–1093. https://doi.org/10.1177/17470218211045987

Elliott, D. B., Sanderson, K., & Conkey, A. (1990). The reliability of the Pelli-Robson contrast sensitivity chart. *Ophthalmic and Physiological Optics, 10*, 21–24. https://doi.org/10.1111/j.1475-1313.1990.tb01100.x

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30*, 3–20. https://doi.org/10.1023/A:1005290706460

Ferris, F. L., III, & Bailey, I. (1996). Standardizing the measurement of visual acuity for clinical research studies: Guidelines from the Eye Care Technology Forum. *Ophthalmology, 103*, 181–182. https://doi.org/10.1016/S0161-6420(96)30742-2

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Futrell, R., & Gibson, E. (2016). L2 processing as noisy channel language comprehension. *Bilingualism: Language and Cognition, 20*, 683–684. https://doi.org/10.1017/S1366728916001061

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences, 110*, 8051–8056. https://doi.org/10.1073/pnas.1216438110

Gibson, E., Sandberg, C., Fedorenko, E., Bergen, L., & Kiran, S. (2015). A rational inference approach to aphasic language comprehension. *Aphasiology, 30*, 1341–1360. https://doi.org/10.1080/02687038.2015.1111994

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science, 28*, 703–712. https://doi.org/10.1177/0956797617690277

Hartshorne, J. K.., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the lifespan. *Psychological Science*, *26*, 433-443. https://doi.org/10.1177/0956797614567339.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystalized intelligence. *Acta Psychologica, 26*, 107–129. https://doi.org/10.1016/0001-6918(67)90011-X

Huang, K.-J., & Staub, A. (2021a). Why do readers fail to notice word transpositions, omission, and repetitions? A review of recent evidence and theory. *Language and Linguistics Compass, 15*, e12434.

Huang, K-J., & Staub, A. (2021b). Using eye tracking to investigate failure to notice word transpositions in reading. *Cognition*, *216*. Advanced online publication. https://doi.org/10.1016/j.cognition.2021.104846.

Kaufman, A. S., Salthouse, T. A., Scheiber, C., & Chen, H. (2016). Age differences and educational attainment across the life span on three generations of Wechsler adult scales. *Journal of Psychoeducational Assessment, 34*, 421–441. https://doi.org/10.1177/0734282915619091

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods, 15*, 722–752. https://doi.org/10.1177/1094428112457829

Levy, R. (2008). A noisy-channel model of human sentence comprehension under uncertain input. In *Proceeding of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 234–243). Stroudsburg, PA: Association for Computational Linguistics.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences, 106*, 21086–21090. https://doi.org/10.1073/pnas.0907664106

Liu, Y., Ryskin, R., Futrell, R., & Gibson, E.. Structural frequency effects in noisy-channel comprehension [Conference presentation]. https://tedlab.mit.edu/tedlab_website/researchpapers/Liu_Ryskin_Futrell_Gibson_Penn_Ling_2021.pdf.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4*, 6–14. https://doi.org/10.1016/S1364-6613(99)01418-7

MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press.

McGowan, V. A., & Reichle, E. D. (2018). The "risky" reading strategy revisited: New simulations using E-Z Reader. *Quarterly Journal of Experimental Psychology, 71*, 179–189. https://doi.org/10.1080/17470218.2017.1307424

Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition, 30*, 551–561. https://doi.org/10.3758/BF03194956

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the selective reanalysis hypothesis. *Journal of Memory and Language, 59*, 266–293. https://doi.org/10.1016/j.jml.2008.06.002

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*, 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundation ideas—Part II. *Language and Linguistics Compass, 10*, 591–613. https://doi.org/10.1111/lnc3.12207

Owsley, C. (2011). Aging and vision. *Vision Research, 51*, 1610–1622. https://doi.org/10.1016/j.visres.2010.10.020

Paape, D., Vasishth, S., & Engbert, R. (2022). Does local coherence lead to targeted regressions and illusions of grammaticality? *Open Mind, 5*, 42–58. https://doi.org/10.1162/opmi_a_00041

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.Rproject.org/.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging, 21*, 448–465. https://doi.org/10.1037/0882-7974.21.3.448

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition, 181*, 141–150. https://doi.org/10.1016/j.cognition.2018.08.018

Ryskin, R., Stearns, L., Bergen, L., Eddy, M., Fedorenko, E., & Gibson, E. (2021). An ERP index of real-time error correction within a noisy-channel framework of human communication. *Neuropsychologia, 158*, Article 107855. https://doi.org/10.1016/j.neuropsychologia.2021.107855

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society, 16*, 754–760. https://doi.org/10.1017/S1355617710000706

Schad, D., Vasishth, V., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a prior contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110*, Article 104038. https://doi.org/10.1016/j.jml.2019.104038

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

Staub, A., Dodge, S., & Cohen, A. L. (2019). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review, 26*, 340–346. https://doi.org/10.3758/s13423-018-1492-z

Sturt, P., & Kwon, N. (2018). Processing information during regressions: An application of the reverse boundary-change paradigm. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01630

Van Gompel, Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language, 45*, 225–258. https://doi.org/10.1006/jmla.2001.2773

Von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language, 65*, 109–127. https://doi.org/10.1016/j.jml.2011.02.004

Warren, T., Dickey, M. W., & Liburd, T. J. (2017). A rational inference approach to group and individual-level sentence comprehension performance in aphasia. *Cortex, 92*, 19–31. https://doi.org/10.1016/j.cortex.2017.02.015

Wechsler, D. (2008). *Wechsler Adults Intelligence Scale – Fourth Edition (WAIS-IV)*. APA PsycTests. https://doi.org/10.1037/t15169-000.

Zhang, J., Warrington, K. L., Li, L., Pagán, A., Paterson, K. B., White, S. J., & McGowan, V. A. (2022). Are older adults more risky readers? Evidence from meta-analysis. *Psychology and Aging, 37*, 239–259. https://doi.org/10.1037/pag0000522

Zhang, Y., Ryskin, R., & Gibson, E. (2023). A noisy-channel approach to depth-charge illusions. *Cognition, 232*, Article 105346. https://doi.org/10.1016/j.cognition.2022.105346