

THIS IS AN AUTHOR ACCEPTED MANUSCRIPT OF A PAPER SET TO APPEAR IN THE *QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY*. THIS IS NOT THE VERSION OF RECORD AND MAY DIFFER FROM THE FINAL PUBLISHED VERSION.

No Evidence of Word-Level Uncertainty in Younger and Older Adults in Self-Paced Reading

Michael G. Cutter¹, Kevin B. Paterson², & Ruth Filik¹

¹ University of Nottingham

² University of Leicester

Word count: 5084 (Excluding References, Footnotes, Tables & Figures)

Author Note

Michael G. Cutter, Ruth Filik, School of Psychology, University of Nottingham, Nottingham, United Kingdom; Kevin B. Paterson, Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, United Kingdom.

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

Correspondence regarding this article should be addressed to Michael G. Cutter, School of Psychology, University of Nottingham, University Park Campus, Nottingham, NG7 2RD. E-mail: michael.cutter@nottingham.ac.uk.

Declarations of interest: none

Abstract

In a self-paced reading study, we investigated whether older adults maintain a greater level of uncertainty about the identity of words in a sentence than younger adults, potentially due to deficits in visuo-perceptual processing of high-spatial frequencies associated with normal aging. In the experiment, 60 older adults and 60 younger adults read sentences in which an early preposition was either perceptually confusable with another word (*at*; confusable with *as*) or not (*toward*), and in which the reading of a subsequent ambiguous verb (e.g. *tossed*) should be affected by the confusability of the preposition, while the reading of an unambiguous verb (e.g. *thrown*) should not be. This design replicated that of an earlier study which found evidence in favour of participants maintaining uncertainty about the confusable preposition in go-past times during natural reading (Levy et al., 2009). However, in our study there was no evidence that either younger or older adults maintained uncertainty about the identity of the perceptually confusable preposition, such that there was no interaction between the preposition's form and subsequent verb ambiguity in self-paced reading times, although we did observe a main effect of verb ambiguity. This represents a failure to replicate the effect observed by Levy et al. when using a different experimental paradigm, and we consider potential causes of our findings at both a methodological and theoretical level.

Keywords; noisy-channel processing; reading; aging; sentence processing.

Aging is typically accompanied by declines in various physiological and psychological functions which are important to processing written language (Salthouse, 2010). For example, older adults typically have poorer visual functioning than younger adults (Owsley, 2011), including a reduced ability to process information from high-spatial frequencies. High-spatial frequencies are important for identifying written words, due to fine-grained letter features being represented by these spatial frequencies, and these features being vital in discriminating between visually similar words (e.g. *at* vs. *as*). Indeed, prior research has shown that older adults struggle more when processing text in which only high-frequency information is available (Jordan et al., 2014). Consequently, older readers may maintain more uncertainty about the identity of words within a sentence than younger readers, which may have downstream consequences for syntactic processing, especially according to “noisy-channel” sentence processing models. We test this proposition in the current study.

Some recent approaches to language processing take the position that readers maintain uncertainty about the identity and presence of prior words in a sentence (Gibson et al., 2013; Levy, 2008; Levy et al., 2009; Ryskin et al., 2018). In these noisy-channel approaches, it is assumed that language processing takes account of the fact that perceptual input is noisy and unlikely to perfectly represent text on the page. Thus, rather than identifying words in an all-or-nothing manner, readers maintain a level of uncertainty about whether a perceived word is actually that word, or if it may be a visually similar word. For example, upon encountering the word *at*, the average reader may assume a probability of 0.85 that the word is *at*, while assigning probabilities of, say, 0.10 and 0.05 to it being *as* or *and*. Furthermore, readers may maintain uncertainty about whether certain words appeared in the prior context at all, such that readers may maintain the possibility that *and* appeared in a sentence in which it was absent, and that they simply missed this word. Another key point of this model is that readers refine beliefs about earlier words’ identities and presence as they

gather additional information from later in the sentence. For example, later words may be less compatible with some possibilities for an earlier word's identity than others, resulting in reduced probability of one possibility and increased probability of other possibilities through Bayesian inference. Importantly, the extent to which new words alter the probability distribution for prior words predicts processing difficulty for each new word, with this difficulty being reflected in reading behaviour.

In a key test of Levy's (2008) noisy-channel account, Levy et al. (2009) conducted an eye-movement study in which participants read sentences such as 1a-d.

1a) The coach smiled at the player tossed the frisbee.

1b) The coach smiled at the player thrown the frisbee.

1c) The coach smiled toward the player tossed the frisbee.

1d) The coach smiled toward the player thrown the frisbee.

Sentences 1a and b contain the word *at*, which, as discussed, readers may assign some probability to being *as* or *and*. In 1c and d *at* is replaced by *toward*, which fulfils the same syntactic function, without being easily confusable with other words. Sentences 1a and c contain the word *tossed* which can be treated as either a past participle or finite verb, while *thrown* in 1b and d can only be treated as a past participle. Crucial to Levy et al.'s argument is the fact that, probabilistically, a finite verb – which *tossed* can be treated as – is more likely to appear in a version of 1a in which *at* is substituted for *as* or *and*, resulting in, for example, “*The coach smiled as the player tossed the frisbee*” as in “*The coach smiled as the player [who] tossed the frisbee [scored a point for his team]*”. Consequently, upon reading *tossed*, readers' beliefs about the earlier preposition's identity will shift, leading to observable processing difficulty. Such a probability shift does not occur when readers encounter *thrown*,

due to *thrown* being impossible to treat finitely, with “*The coach smiled as the player [who] thrown the frisbee...*” being ungrammatical.

When *at* is replaced by *toward*, the situation differs. There is no word identity uncertainty, since *toward* has no perceptually similar neighbours with which it can be substituted. Thus, there is no difficulty driven by uncertainty about the preposition at one verb relative to the other. However, this does not equate to no uncertainty about context preceding the verb – within noisy-channel accounts readers also assign a probability that a word is missing from the preceding context, with the addition of this word causing verb ambiguity. For instance, including *and* after *player* in the sentence also results in a case where *tossed* can be treated finitely. Consequently, readers may still experience processing difficulty at the ambiguous verb when it is preceded by a non-confusable preposition; however, it is crucial to Levy et al.’s argument that there is extra uncertainty and difficulty when the preposition is confusable. Levy et al. demonstrated that readers exhibited greater reading difficulty upon encountering the ambiguous verb in 1a relative to the other conditions, such that they took longer to move their gaze beyond this word, and were more likely to make regressions to earlier words and make comprehension errors.

Returning to aging’s effect upon vision, a reduced ability to process high-spatial frequencies should result in greater perceptual noise, and thus greater uncertainty about word identity. To return to the word *at*, if older readers maintain greater uncertainty about word identities they may hypothetically assign probabilities of, say, 0.76, 0.16, and 0.08 to the word being *at*, *as*, or *and*, with this representing more uncertainty than in younger readers (hypothetically, 0.85, 0.10, and 0.05). Increased uncertainty about the identity of *at* means information extracted from later in the sentence may exert more influence on older readers’ beliefs about the identity of *at*, resulting in larger updates in beliefs than in readers who were originally more certain about the preposition’s identity, due to the nature of Bayesian

inference (i.e. flatter priors allow new information to more strongly influence posterior beliefs). Consequently, older readers may experience greater processing difficulty at this point in the sentence due to greater uncertainty about *at*. Specifically, reading times for younger and older adults should be longer upon encountering *tossed* versus *thrown*, with this effect being particularly pronounced when *at* appeared earlier in the sentence. Furthermore, this effect may be larger for older adults. We address this issue by presenting older and younger adults with the sentences used by Levy et al. in a self-paced reading study, with the goal of determining whether older readers maintain more uncertainty about word identity. It is worth emphasising that our study was conducted using self-paced reading methodology as opposed to eye-tracking during reading. Thus, beyond the question of whether older adults maintain uncertainty about word identity during reading differentially to younger adults, it will also be interesting to consider if readers in general maintain uncertainty about word identity when they are constrained to only being able to move forwards through the text, and not regress backwards.

Method

Participants

We collected and analysed data from 60 young (age range 18-29 years, age mean=23 years; 42 females, 1 non-binary; mean years of education=14.5; mean hours spent reading per week=15) and 60 older adults (range 65-80 years, mean=70 years; 35 female; mean years of education=13.6; mean hours spent reading per week=19). Each age group sample was 1.5 times the size of Levy et al.'s full sample, and our full sample was three times as large. Seven additional participants were tested but excluded for varying reasons. One young adult was excluded for accuracy below 75% on easy comprehension questions about filler stimuli. Secondly, we fit a regression model to each participant's reading data, in which region length

(rounded to the nearest 10 characters) was treated as a predictor variable. If there was no significant effect of region length on a participant's reading time we assumed low task engagement, and excluded their data; this affected three young and one older participant. Finally, two young adults were excluded due to technical issues resulting in over-recruitment for two counterbalancing conditions –we excluded the final participant to contribute data to each condition. Participants were recruited via Prolific academic, with younger and older adults receiving rewards of £5.75 and £7.50, respectively.¹

Materials and Design

We presented readers with 24 items similar to sentences 1a-d. These were counterbalanced across participants, with each item appearing in each condition for 25% of participants, and each participant reading 25% of items per condition. These items were the same as those used by Levy et al., although some required localisation for the British-English speaking participants we tested. For each item we constructed a comprehension question similar to those used by Levy et al; 20 probed relative clause verb comprehension, while four probed main verb understanding. The unambiguous verb was on average 5.7 letters long with a zipf frequency in the British National Corpus of 2.25, while the ambiguous verb was on average 6.4 letters long with a mean zipf frequency of 2.55. The difference in length was significant when tested with a paired t -test ($t(23) = 2.27, p = .033$), while the difference in frequency was not ($t(23) = 1.50, p = .148$). As such, we controlled for length statistically in the main reading time analyses presented below, and we controlled for frequency in an additional supplementary analysis. For the ambiguous verbs, there was a clear preference (i.e. over 60% of usages) for the simple past tense form for 13 items, a clear preference for the past participle for five items, and no clear preference for the remaining six items.

¹ This discrepancy in payment was due to older adults generally taking longer than younger adults to complete the experiment.

These materials were presented alongside 24 items from another experiment – 12 including garden-path constructions – and 48 fillers with simple structures including a word frequency manipulation. This frequency manipulation was included in the fillers to ensure that the implementation of self-paced reading we used and online participant sampling could reproduce one of the more replicable effects in the literature. This was successful, with significant frequency effects observed on target and post-target words. In our experiment, 36 out of 103 items (35%) consisted of syntactically complex sentences (i.e. the Levy et al. stimuli combined with the garden-path stimuli), a similar percentage as in the original Levy et al. study (35% syntactically complex sentences, or 24 out of 68 items).

Procedure

Our procedure was approved by the University of Nottingham's School of Psychology Ethics Committee [F1258]. We presented our study using Gorilla.Sc, a web-based client for conducting behavioural experiments (see Anwyl-Irvine et al., 2020) with high precision in measuring reaction times. For example, Bridges et al. (2020) found that in recorded reaction times there was a standard deviation of only 5.3ms from actual reaction time, even in non-optimal set-ups.

After providing informed consent, participants completed several tasks. First, they provided demographic information (i.e. age, years of education, hours spent reading per week) and completed a bot-check. Next, they completed the self-paced reading task. We implemented our phrase-by-phrase non-cumulative self-paced reading task using Gorilla's Reading Zone feature. In this implementation, participants were initially presented with the sentence's first region, while the rest of the sentence remained masked by black highlighting. Each time participants hit the spacebar the next region was revealed, with prior text being masked.

We divided our stimuli into five presentation regions (see Figure 1). Region Three was the target region, consisting of the reduced relative verb where we expected to observe effects of our manipulation. Region Four comprised the following noun-phrase and was examined for spill-over effects. Each experimental sentence was followed by a yes-no comprehension question.

R1 R2 R3 R4 R5
The coach| smiled at the player| tossed| the frisbee| by the opposing team.

Figure 1. An example of how our stimuli were divided into presentation regions in self-paced reading. The vertical pipes represent the point at which one region ends and another begins.

After the self-paced reading task, participants completed a digit-span and Stroop task. These were included in relation to another experiment within the reading task. Participants were debriefed, and given an opportunity to provide information about any technical issues they experienced.

Results

We analysed our data in the R statistical computing environment (R Core Team, 2020), using the *brms* package (Bürkner, 2017; version 2.5.0). Before analysis, we log-transformed reading times because of rightward skew, and removed observations over three standard deviations from the grand mean. We used reading time as a dependent variable in lognormal Bayesian mixed-models (see Nicenboim & Vasisht, 2016, for an introduction to using Bayesian methods in linguistic research), with age group (younger vs. older), preposition (*towards* vs. *at*), and verb ambiguity (ambiguous vs. unambiguous) as predictor variables, alongside two-way and three-way interactions between these variables. For the

target region, verb length was included as a predictor variable.² Participants and items were treated as random effects, with maximal random structures. For predictor variables we used weakly informative priors of $Normal(0, 1)$ which accounted for data being log-transformed, with a regularization of 2 on the covariance matrix of random effects. Separate models with four chains of 8000 iterations each – with the first 1000 chains used as warmup – were run for our target region, spill-over region, and on comprehension rates for questions probing reduced relative verb interpretation. We report the model's median effect estimates, upper and lower 95% credible intervals (CrI), and the probability of effects being in a certain direction ($P(\hat{b} > 0)$). Mean reading times per condition for target and spill-over regions are displayed in Table 1, alongside comprehension rates; Table 2 presents model estimates for the target verb.

On the target verb there was a main effect of age group, with older adults taking longer to read than younger adults. The hypothesised three-way interaction between age group, verb ambiguity, and preposition was not observed. More saliently, we failed to observe an interaction between the preposition's form and verb ambiguity, representing a failure to find the same evidence for word-level uncertainty which Levy et al. observed in go-past times, contradicting the idea that readers maintained uncertainty about the identity of *at* (see below for further analysis pertinent to this point). However, there was a verb ambiguity main effect, with participants spending longer reading ambiguous than unambiguous verbs, regardless of the preceding preposition, demonstrating that we could detect syntactic difficulty effects in our experiment. While this effect was numerically larger in older (123ms)

² In the main analysis presented here, we did not include the target verbs' zipf frequency as a predictor variable, since the difference in frequency across conditions was not significant. However, due to a comment made by a reviewer of an earlier versions of our manuscript we did include verb frequency as a predictor variable in an additional analysis. The inclusion of this predictor variable made no notable differences to our analysis, such that evidence for the main effect of verb ambiguity was unaffected, while the interaction between verb ambiguity and preposition remained absent.

than younger adults (67ms), there was no statistical evidence for an interaction between age and verb ambiguity.

Table 1

Mean reading times (and standard errors) in milliseconds per condition in each region, and comprehension rates in proportion correct.

| | Younger adults | | | | Older adults | | | |
|-------------------|----------------|-------------|-------------------|-------------------|--------------|-------------|-------------------|-------------------|
| | At-Thrown | At-Tossed | Toward- Thrown | Toward- Tossed | At-Thrown | At-Tossed | Toward- Thrown | Toward- Tossed |
| Target Region | 731 (24) | 791 (29) | 739 (26) | 813 (31) | 1080 (29) | 1221 (38) | 1084 (30) | 1189 (33) |
| Spill-over Region | 772 (28) | 773 (29) | 732 (28) | 713 (27) | 1055 (35) | 1019 (33) | 1060 (34) | 1015 (30) |
| Comprehension | 0.63 (0.02) | 0.64 (0.02) | 0.71 (0.02) | 0.62 (0.02) | 0.73 (0.02) | 0.71 (0.02) | 0.77 (0.02) | 0.71 (0.02) |

In the spill-over region older adults had longer reading times ($b=-0.34$, CrI[-0.44,-0.23], $P(\hat{b}>0)=1$), with the interaction between preposition and verb being absent ($b=0.00$, CrI[-0.06,0.06], $P(\hat{b}>0)=0.53$). There were no other effects in this region.

In terms of comprehension, older adults performed better than younger adults ($b=-0.47$, CrI[-0.86,-0.08], $P(\hat{b}<0)=.99$). However, there was little evidence of misunderstandings increasing due to uncertainty about the preposition's identity, with comprehension highest in sentences featuring *toward* and an unambiguous verb, and the three remaining conditions having similar comprehension levels. Estimates for the remaining effects in our statistical model contained 0 within their credible intervals, suggesting little systematic influence of these factors on comprehension. Furthermore, any trend towards an interaction between verb ambiguity and preposition was in the direction opposite to that observed by Levy et al., with verb ambiguity affecting comprehension more given the preposition *toward* than *at*.

Table 2

Population-level effects from the Bayesian model for the target word *tossed/thrown*, including median effect estimate, lower and upper credible intervals, and the probability that the effect was in the more likely direction.

| | Estimate | L-95% CI | U-95% CI | $P(\hat{b} > / < 0)$ |
|----------------|----------|----------|----------|----------------------|
| Intercept | 6.73 | 6.68 | 6.78 | 1 |
| Age | -0.42 | -0.52 | -0.33 | 1 |
| Preposition | 0.00 | -0.03 | 0.03 | 0.52 |
| Verb Ambiguity | -0.06 | -0.10 | -0.01 | 0.99 |
| Verb Length | 0.03 | 0.01 | 0.05 | 0.99 |
| Age*Prep | 0.01 | -0.06 | 0.07 | 0.57 |
| Age*Verb | 0.03 | -0.04 | 0.09 | 0.80 |
| Prep*Verb | 0.00 | -0.06 | 0.07 | 0.56 |
| Age*Prep*Verb | -0.02 | -0.14 | 0.10 | 0.62 |

Note. We consider there to be unequivocal evidence for effects when there is a probability of above .975 of the effect being in a certain direction and the credible interval does not contain the value of 0.

To follow-up on aspects of our data, we calculated Bayes factors (see Lee & Wagenmakers, 2013) to determine the extent to which our target region data represented evidence for or against certain effects. Bayes factors provide a ratio of evidence for one statistical model versus another, such that when comparing Model 1 with Model 0 a Bayes factor above 1 suggests the data supports Model 1 while a Bayes factor below 1 supports

Model 0. Evidence is considered sufficient to support one model over another when the ratio is greater than 3 (i.e. values above 3 provide sufficient evidence for Model 1; values below 0.33 provide sufficient evidence for Model 0; values in-between are considered merely anecdotal evidence).

The main effect of interest was the interaction between verb ambiguity and preposition. This interaction represented Levy et al.'s evidence for word level uncertainty, and so it is important to ascertain whether our data represents evidence against this interaction. We also calculated Bayes factors for the main effect of verb ambiguity, to ensure that our approach to calculating Bayes factors and the priors we used did not overly favour the null effect.

To calculate Bayes factors for the interaction between verb ambiguity and preposition we compared a model including the interaction to a model including only additive effects of these variables. To calculate Bayes factors for the verb ambiguity main effect we compared a model including this main effect to one not including the effect.³ Models were run with 20,000 iterations each. Bayes factors were calculated using bridge sampling, with this repeated 20 times due to slight variations in output. We report the average of these 20 repetitions.

Due to sensitivity of Bayes factors to the prior we repeated the above process with priors of $normal(0,1)$, $normal(0,.1)$, and $normal(0,.05)$. We used the two latter more informative priors since uninformative priors overly favour the null hypothesis, due to assigning probability to unrealistically large effect sizes. Specifically, a prior for an effect size of $normal(0,1)$ with a mean reading time of ~955ms assigns a relatively large probability to unrealistically large effect sizes, such that the prior assigns a probability of ~68% that the

³ See associated R scripts at https://osf.io/ynd9s/?view_only=7a697030c13a424781bed08be108f1a6 for full model specification.

effect of a predictor term in the model is between -605ms and +1636ms, a range which we suspect most psycholinguists would agree is unlikely. When the Bayes factor is calculated using this type of prior, the null effect is favoured since the question we are asking is essentially “is it more likely that there is a null effect of this variable or an unrealistically large effect of this variable?” On the other hand, with the prior of *normal*(0,.05) the prior suggests a probability of ~68% that the effect is between -48ms and 47ms, a much more realistic range. Here, the question we are asking when calculating the Bayes factor becomes “is it more likely that there is a null effect of this variable or a reasonably small effect of this variable?”

Bayes factors are displayed in Table 3. There was very strong evidence against the interaction effect with an uninformative prior, moderate evidence against with a more informative prior, and anecdotal evidence against with the most informative prior. For the main effect of verb ambiguity there was anecdotal evidence against with the uninformative prior, but moderate evidence for an effect with the informative prior.

Table 3

Bayes factors for the interaction between preposition form and verb ambiguity and the main effect of verb ambiguity in our target region with priors of varying informativeness.

| | Normal (0,1) | Normal (0,.1) | Normal (0,.05) |
|------------------|--------------|---------------|----------------|
| Preposition*Verb | .029 | .327 | .521 |
| Verb | .475 | 5.08 | 6.95 |

Supplementary analysis

While it has long been standard practice for researchers to log-transform reading time measures in order to make this measure normally distributed for statistical analysis, it has recently been argued that this can be inappropriate, with such transformations suppressing interactions that are actually present in the data (see Lo & Andrews, 2015). This could be considered particularly problematic in a paper such as ours, whereby we have failed to observe an interaction that previously published research suggests should be present. Due to these concerns, we present an additional analysis of the data for the critical verb, in which we instead analyse the raw reaction time data using a linear mixed-model which assumed the distribution of our dependent variable was described by the gamma distribution rather than normal distribution, using an identity link function. We implemented this model in a frequentist framework using the lme4 package (Bates et al., 2015; version 1.1-25) in R.

This analysis suggested a near identical pattern of effects to the lognormal model presented above. Specifically, the model still suggested that there were effects of age ($b = -361$, $SE = 42$, $t = -8.6$, $p < 0.001$), verb ambiguity ($b = -75$, $SE = 24$, $t = -3.2$, $p = 0.001$), and verb length ($b = 18$, $SE = 8$, $t = 2.3$, $p = 0.026$). Most importantly, given concerns about log-transformations suppressing interactive effects, there was still no interaction between preposition form and verb type ($b = 7.6$, $SE = 39$, $t = 0.2$, $p = 0.847$), with the 7.6ms trend predicted by the model actually being representative of preposition form affecting reading times on unambiguous verbs but not ambiguous verbs, a pattern of effects opposite to our hypothesis.

Discussion

We set out to examine whether older adults maintain greater uncertainty about the identity of prior words in a sentence than younger adults. We presented readers with sentences in which the ambiguity of a verb's syntactic function could be influenced by

uncertainty about an earlier preposition's identity. Prior research using eye-tracking methodology has shown that a group of 40 younger adults experienced greater processing difficulty upon encountering the ambiguous verb when it followed a more orthographically confusable preposition, and had lowest comprehension in these conditions (Levy et al., 2009). We hypothesised that difficulty in these conditions may be exacerbated in older adults, due to visual processing deficits (Jordan et al., 2014). Unexpectedly, we failed to find evidence of the effects observed by Levy et al. when we used self-paced reading as a methodology instead of eye-tracking. The key interactive effect was absent in both age groups tested, and participants instead experienced processing difficulty due to verb ambiguity regardless of the preceding preposition. Bayes factors suggested anecdotal evidence against the interactive effect with a highly informative prior and moderate to very strong evidence against it with less informative priors. In contrast, there was moderate evidence for a main verb ambiguity effect with informative priors, and weaker evidence against it with an uninformative prior. Thus, the focus of our discussion will not be age differences in noisy-channel processing and word identity uncertainty, but an assessment of why we did not observe an interaction between preposition and verb ambiguity, while still observing verb ambiguity effects, and the role that methodology may have played in our pattern of findings.

One obvious possible cause for not observing an interaction between preposition form and verb ambiguity is that we used self-paced reading rather than eye-tracking during reading. It could be that, when readers are denied the possibility of moving their eyes backwards in text, they are less likely to maintain word identity uncertainty, due to being unable to collect further perceptual evidence at a later point. Indeed, Levy et al. primarily found evidence for their effects in eye-movement measures accounting for regressions, with a 77ms verb ambiguity effect in go-past times for sentences featuring the word *at*, but no effect for sentences featuring *toward*. Thus, it could be argued that the reason we did not observe

this same interaction is because self-paced reading is simply not a sensitive enough measure to detect effects relating to uncertainty about prior linguistic context, or it could even be the case that readers cease maintaining uncertainty altogether when they are unable to make regressions. This is certainly a fair criticism of the current study. However, there are several aspects of our data which we believe make it worth considering the possibility that our results represent evidence against readers maintaining uncertainty about word identity during reading in general, as opposed to our findings simply being an artifact of self-paced reading methodology.

One aspect of our data that is incompatible with readers simply not maintaining uncertainty during self-paced reading is that— at least within a noisy-channel account— we should have observed a null effect of verb ambiguity, regardless of preposition. Instead, we observed disruptive verb ambiguity effects regardless of preposition. Within the noisy-channel framework, main effects of verb ambiguity occur because readers maintain uncertainty about the presence of certain words earlier in the context, such that *tossed* can be treated finitely if for example the word *and* is inserted after *the player* (i.e. *The coach smiled at/towards the player [and] tossed the frisbee*). There are limits here, such that prior self-paced reading research has shown that verb ambiguity effects disappear when the relativizer *who was* precedes the target verb (i.e. *The coach smiled at the player [who was] tossed/thrown the frisbee*; Tabor et al., 2004), presumably because the transformations necessary to the prior context to treat the verb finitely are relatively low probability (i.e. readers would need to have imagined the word *was*). Within the noisy-channel framework and the current study, it is unclear why effects driven by uncertainty about the presence of a word were still present in self-paced reading, while those due to uncertainty about a word's identity were not. While Levy et al. did not observe a main effect of verb ambiguity in go-

past times, other prior investigations of local coherence effects have found such effects (e.g. Christianson et al., 2017), and Levy et al. observed these effects in gaze durations.

Another aspect of our findings that we do not think can be explained by the use of self-paced reading as a methodology are the null effects in comprehension. Levy et al. found lower comprehension accuracy in sentences including *at* and an ambiguous verb. Within noisy-channel accounts this effect is presumably driven by readers reaching the ambiguous verb, assuming it is finite, and updating beliefs such that *at* is treated as *as* in the final sentence representation, leading to misinterpretations. During natural reading, readers can regress back to *at* to check its identity, with this resulting in beliefs being updated back towards the correct word, due to increased perceptual evidence about word identity. Regressions are impossible in self-paced reading, meaning that participants could not gather more perceptual evidence from the preposition in response to encountering the ambiguous verb. Consequently, it seems that, if anything, the size of this effect on comprehension should have increased in our study due to the paradigm used, as opposed to being entirely absent. This explanation does of course depend on the idea that the reason the effects did not appear in our reading time measure was simply due to this measure only including forward reading times, as opposed to readers stopping maintaining any uncertainty at all during self-paced reading.

There are, broadly speaking, two potential competing sets of implications to take from our study, dependent upon the extent to which our findings are either a) due to self-paced reading methodology or b) representative of more general reading behaviour. If our failure to find the same interaction as Levy et al. is a consequence of self-paced reading, these findings nevertheless have implications for noisy-channel accounts of sentence processing, as elaborated above. To re-iterate, it is unclear why effects attributed to uncertainty about the presence of certain words (i.e. a main verb effect) appear in self-paced reading, while effects

attributed to uncertainty about word identity (i.e. an interaction between preposition and verb) do not. One intriguing possibility is that two somewhat independent mechanisms account for uncertainty about word identity and word presence. For example, uncertainty about word identity could be treated primarily as a perceptual problem, whereby readers maintain uncertainty due to their encoding of the input. Here, being prevented from sampling further perceptual evidence from this word may eliminate the propensity to maintain uncertainty. In contrast, uncertainty about the presence of certain words could in part be treated as a problem in sentence production. Here, readers may assume that certain words were erroneously omitted from the sentence by the sentence producer, with an inability to return to the context not affecting this process. The idea that noisy-channel inferences are driven by potential production errors as opposed to comprehension errors is not new (e.g. see Gibson et al., 2017), although the distinction in the extent to which certain processes drive different types of inferences may be.

Alternatively, it could be that our null interaction is representative of more general reading behaviour, with the effects observed by Levy et al. (2009) being non-replicable. Recall that Levy et al. presented 24 items to 40 participants, representing a relatively small sample. A large-scale replication of this work examining eye movements during reading may be necessary to establish the replicability of Levy et al.'s original finding, or, alternatively, to confirm the null finding observed in the current self-paced reading study. Work with this aim is currently underway (Cutter et al., 2021).

It is also worth considering alternative accounts of the verb ambiguity effects observed in our study, which view these as local coherence effects. For example, Tabor et al. (2004) explained these effects within a self-organised parsing account (SOPARSE; Tabor & Hutchins, 2004). In SOPARSE, each word introduces fragments of syntactic trees, with potential open attachment sites. The tree-fragment introduced by each word will attempt to

attach with the tree-fragments from other words. In our sentences, the tree-fragments for the words in the relative clause can combine into a locally coherent active clause structure which is incompatible with the wider sentence context when the verb is ambiguous (e.g. *the player tossed the frisbee*), while items featuring an unambiguous verb cannot be treated as active clauses (e.g. *the player thrown the frisbee*). In SOPARSE processing difficulty occurs because the globally infelicitous (but locally coherent) active clause analysis of the relative clause competes with the main sentence parse. In other words, disruption is partly due to readers needing to inhibit the active clause reading of the relative clause. Crucially, within this explanation there should be no interaction between verb ambiguity and preposition form, meaning that it can explain our basic findings.

The issue of aging effects on noisy-channel processing is worth briefly returning to, despite being a question we cannot answer in the current paper. We originally hypothesised older adults may experience greater word uncertainty, resulting in greater processing difficulty at the ambiguous verb when it was preceded by *at*. However, it should be noted that older adults may not necessarily experience greater processing difficulty, even if they maintain greater uncertainty about the preposition's identity – for example, increased uncertainty may also affect the certainty with which the verb is identified, thus decreasing the extent to which the syntactic ambiguity of this word alters beliefs about preceding context, contrary to our initial hypothesis. However, precisely how effects of noisy-channel processing might be predicted to change in light of visual and cognitive declines in late adulthood presently is unclear but could provide a valuable framework for investigating reading comprehension across the lifespan.

In closing, we set out to test whether older adults' reduced ability to process high-spatial frequencies would result in them maintaining greater word identity uncertainty than younger adults. Unexpectedly, our data showed no evidence of either age group maintaining

uncertainty about word identity. It may be that a future large-scale eye-movement study is necessary to test whether this discrepancy was due to methodology, or whether Levy et al.'s sample did not provide an accurate estimation of the effect size as it exists in the population. In either scenario, the current work presents important implications for noisy-channel approaches to language process.

Declaration of Conflicting Interests

The authors declare that there are no conflicting interests.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Leverhulme Trust [grant number RPG-2019-051].

References

- Anwyl-Irvine, A. L., Massoné, J., Flitton, A., Kirkham, N. & Evershed, J. K. (2020). Gorilla in our midst: An online behavioural experiment builder. *Behavior Research Methods*, 52, 388-407. <https://doi.org/10.3758/s13428-019-01237-x>
- Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48. doi:10.18637/jss.v067.i01.
- Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8:e9414. <https://doi.org/10.7717/peerj.9414>
- Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1-28. doi:10.18637/jss.v080.i01
- 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. <http://dx.doi.org/10.1080/17470218.2016.1186200>
- Cutter, M. G., Filik, R., & Paterson, K. B. (2021). Do Readers Maintain Word-Level Uncertainty during Reading? A Pre-Registered Replication Study. OSF. <https://osf.io/vw3fd/>
- 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., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science*, 28, 703-712. <https://doi.org/10.1177/0956797617690277>

- Jordan, T. R., McGowan, V. A., & Paterson, K. B. (2014). Reading with filtered fixation: Adult age differences in the effectiveness of low-level properties of text within central vision. *Psychology and Aging, 29*, 229-235. <https://doi.org/10.1037/a0035948>
- Lee, M.D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modelling: A practical course*. Cambridge University Press.
- 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).
- 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>
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1171.
<https://doi.org/10.3389/fpsyg.2015.01171>
- 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>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.Rproject.org/>.

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>

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, *16*, 754-760.

<https://doi.org/10.1017/S1355617710000706>

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*, 355-370.

<https://doi.org/10.1016/j.jml.2004.01.001>

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

30, 431-450. <https://doi.org/10.1037/0278-7393.30.2.431>