Executive Resources Shape the Effects of Language Predictability Across the Adult Lifespan

Merle Schuckart[1,2]*, Sandra Martin[3]*, Sarah Tune[1,2], Lea-Maria Schmitt[4], Gesa Hartwigsen[3,5]⋆ & Jonas Obleser[1,2]⋆

[1] Department of Psychology, University of Lübeck, Germany
[2] Center of Brain, Behavior and Metabolism, University of Lübeck, Germany
[3] Research Group Cognition and Plasticity, Max Planck Institute for Human Cognitive and Brain Sciences, Germany
[4] Donders Institute for Brain, Cognition and Behaviour, Radboud University, The Netherlands
[5] Wilhelm Wundt Institute for Psychology, Leipzig University, Germany
* Shared first authorship
⋆Shared last authorship

ABSTRACT

Humans predict upcoming language effortlessly, but the cognitive costs of these predictions remain debated. In this study, we demonstrate that limiting executive resources through a concurrent task reduces the effect of language predictability on reading times. Participants (N=175, replication N=96) read short articles presented word-by-word while completing a secondary font colour n-back task, thus varying cognitive demand. Language predictability was indexed by word surprisal as derived from a generative pre-trained transformer model (GPT-2). Our results show that predictability effects increase with age (18–85 years), and across both samples, higher cognitive load consistently diminished the impact of word predictability on reading speed. These findings highlight the resource-dependent nature of predictive mechanisms in language comprehension across the lifespan. We conclude that linguistic predictions rely on executive resources, which helps resolve the debate about the cognitive cost of predictive language processing.

INTRODUCTION

In everyday life, we constantly rely on our ability to swiftly yet accurately process linguistic input. When reading a book, watching television, or navigating a car through busy traffic while following instructions, language prediction is considered a catalyst that enhances the efficiency of linguistic processing (Clark, 2013; Pickering & Garrod, 2007; Onnis et al., 2022; Rao & Ballard, 1999; Ryskin & Nieuwland, 2023). However, due to the inherent flexibility and richness of natural language, upcoming words can rarely be predicted from context with complete certainty (Rubenstein & Aborn, 1958). Instead, linguistic features are thought to be pre-activated broadly rather than following an all-or-nothing principle, as there is evidence for predictability effects even for moderately- or low-restraint contexts across the entire predictability span (Boston et al., 2008; Roland et al., 2012; Schmitt et al., 2021; Smith & Levy, 2013). This generation of graded predictions is sometimes described as being passive and cost-free (Luke & Christianson, 2016). However, it is still under debate whether maintaining such an elaborate process really incurs no cognitive cost.

Graded language predictions necessitate the active generation of hypotheses on upcoming words as well as the integration of prediction errors to inform future predictions (Clark, 2013; Ryskin & Nieuwland, 2023). Supporting this, recent evidence suggests that language predictions may indeed impose processing demands: Shain et al. (2024) found that reading times increase with decreasing word predictability, with even small drops in predictability of highly expected words leading to significant processing costs. These findings suggest that language predictions are not entirely automatic or effortless. This aligns with numerous neuroimaging studies arguing for a strong interaction between language-specific and domain-general executive brain regions (Geranmayeh et al., 2017; Martin et al., 2022; Sliwinska et al., 2017; Wingfield & Grossman, 2006), particularly in situations that are cognitively demanding (Erb et al., 2013; Peele, 2018; Vaden et al., 2013; Vaden et al., 2015; Vaden et al., 2016). However, to date, it is unclear to which extent language predictions tax executive resources (MacGregor et al., 2022; Shain et al., 2022; Xie, Brodbeck & Chandrasekaran, 2023; Ryskin & Nieuwland, 2023).

An interesting test case for the dynamic interplay between the two systems is cognitive ageing, as advancing age has been shown to be associated with sensory and executive decline (Idowu & Szameitat, 2023; Salthouse, Atkinson & Berish, 2003; Verhaeghen et al., 2003). In the current study, we thus ask: How does language prediction change when executive resources are limited – both intrinsically due to advanced age, and extrinsically through increased task demands?

The age-related change in cognitive resources is reflected by longer linguistic processing times, especially in situations with high cognitive effort, such as dual-task processing (Liu et al., 2016; Smiler et al., 2003). However, previous research presents conflicting results on how cognitive ageing affects the use of linguistic predictions during language comprehension. Behavioural studies suggest a decline in using context to make semantic predictions with age (e.g., Häuser et al., 2019). Yet, EEG studies present a more nuanced picture, with some findings indicating heightened sensitivity to unexpected information in older adults (Cheimariou et al., 2019), while others report no significant age-related differences in neural responses (Dave et al., 2018; Payne & Federmeier, 2018). It is further unclear how the use of language predictions across the adult lifespan might be affected by increasing cognitive demands due to dual-task processes. According to the *compensation-related utilisation of neural circuits* (CRUNCH) hypothesis (Reuter-Lorenz & Cappell, 2008), older adults might earlier reach a point where their cognitive load capacity is fully exhausted than young adults, leading to a performance decline. If language prediction draws on executive resources, its effect on reading times might thus diminish with increasing cognitive load due to shared cognitive resources. However, this effect might re-emerge once capacity limits are reached, causing tasks to be processed sequentially, which would result in an increased effect of language prediction paralleled by slower performance.

Here, we explored the role of executive control in language prediction in two large cohorts across the adult lifespan. Using a novel dual-task paradigm that couples natural reading with an n-back task as a proxy of executive functioning, we tested the following hypotheses: First, both increased cognitive load and reduced word predictability should be reflected by longer reading times (Fig. 1a and b), alongside a decrease in text comprehension and n-back task performance. Second, we expected that the formation of language predictions should be contingent upon the availability of executive resources. Accordingly, a gradual limitation of these resources due to increased task demands should result in diminished effects of language predictability on reading times (Fig. 1c). Third, we explored how these previously described effects would be modulated by age: Expanding upon our hypothesis that language predictions should be impaired if executive resources are limited, we hypothesised that with increasing age and thus decreasing executive resources (Idowu & Szameitat, 2023; Salthouse, Atkinson & Berish, 2003), individuals should rely less on language predictions. This should be reflected by diminished predictability effects (Fig. 1d, bottom). However, given the presumed stability of language comprehension across the lifespan (Shafto & Tyler, 2014), we also considered the opposite case, namely, that older adults might rely more heavily on language predictions, thus fully compensating for any

impairments in reading comprehension caused by their sensory and executive decline (Idowu & Szameitat, 2023; Salthouse, Atkinson & Berish, 2003; Verhaeghen et al., 2003). In this case, we should see strong surprisal effects independent of age (Fig. 1d, top).

Our work helps resolve the ongoing discussion on the role of executive resources in language predictions with three key findings: First, across our large age range, we found a general increase in reading time with both low language predictability and high task demands. Second, higher task demands reduce the influence of language predictability on reading time, indicating that linguistic prediction relies on executive control resources. Third, predictability had a more pronounced effect on reading efficiency in older adults compared to younger ones. These findings highlight the dynamic interaction between language predictability and executive resources.

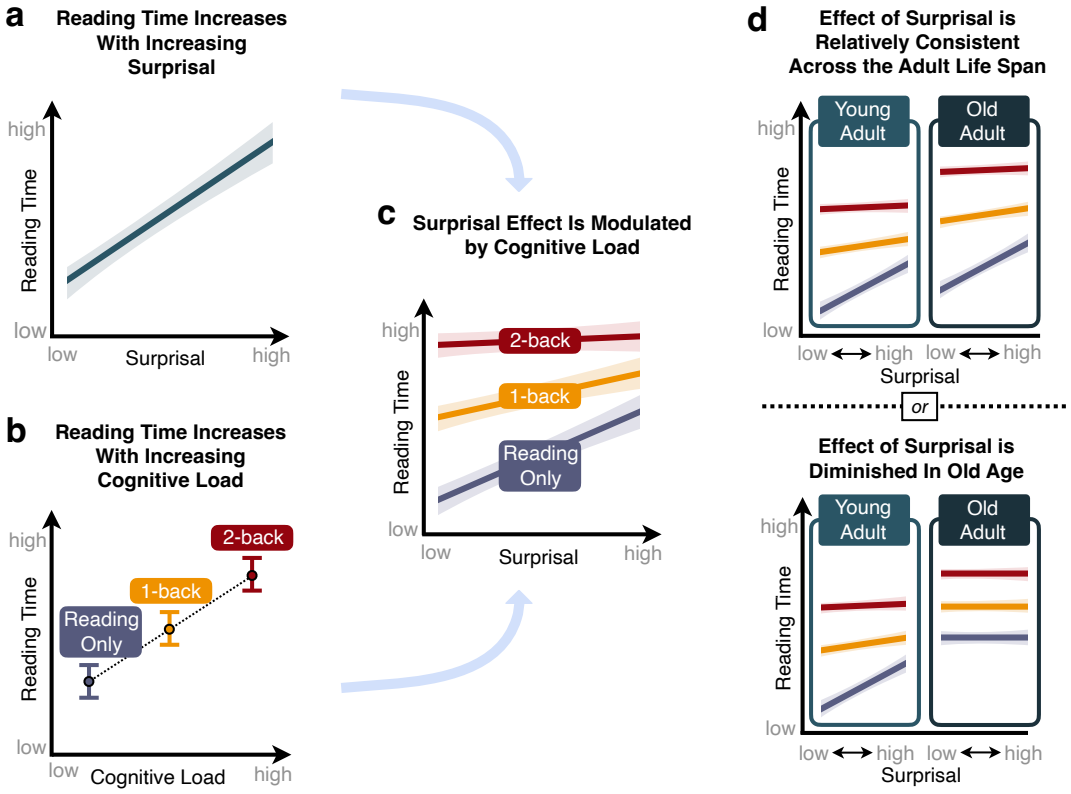

Figure 1. Visualisation of hypotheses. We expected main effects on reading time of **(a)** surprisal and **(b)** cognitive load, as well as **(c)** an interaction of surprisal and cognitive load. Additionally, **(d)** we explored how these effects are modulated by age. Here, two alternative scenarios seemed equally plausible: The effect of surprisal could either be consistent across age groups (**d**, top panel; no moderation by age), or language prediction deteriorates with declining executive resources in old age, resulting in the absence of a surprisal effect in older age groups (**d**, bottom panel).

RESULTS

We report data from 175 participants (M = 44.9 ± 17.9, 18-85 years, 51% female) who were either tested online (N_{online} = 80) or in the laboratory ($N_{laboratory}$ = 95) in one session. Moreover, we conducted an internal, pre-registered replication study involving another 96 participants (M = 39.8 ± 14.0, 18-70 years, 51% female) tested online to confirm the reliability of our findings (see Fig. S1 for age distributions of all samples). During the experiment, participants engaged in a self-paced reading task. They read 300-word newspaper articles, presented word-by-word in various font colours (Fig. 2a). The task was either performed in isolation (*Reading Only*) or paired with a competing n-back task on the words' font colour (*1-back* and *2-back Dual Task*). Participants were instructed to read the texts carefully, as content-related multiple-choice questions were asked after each text (see *Methods* for details).

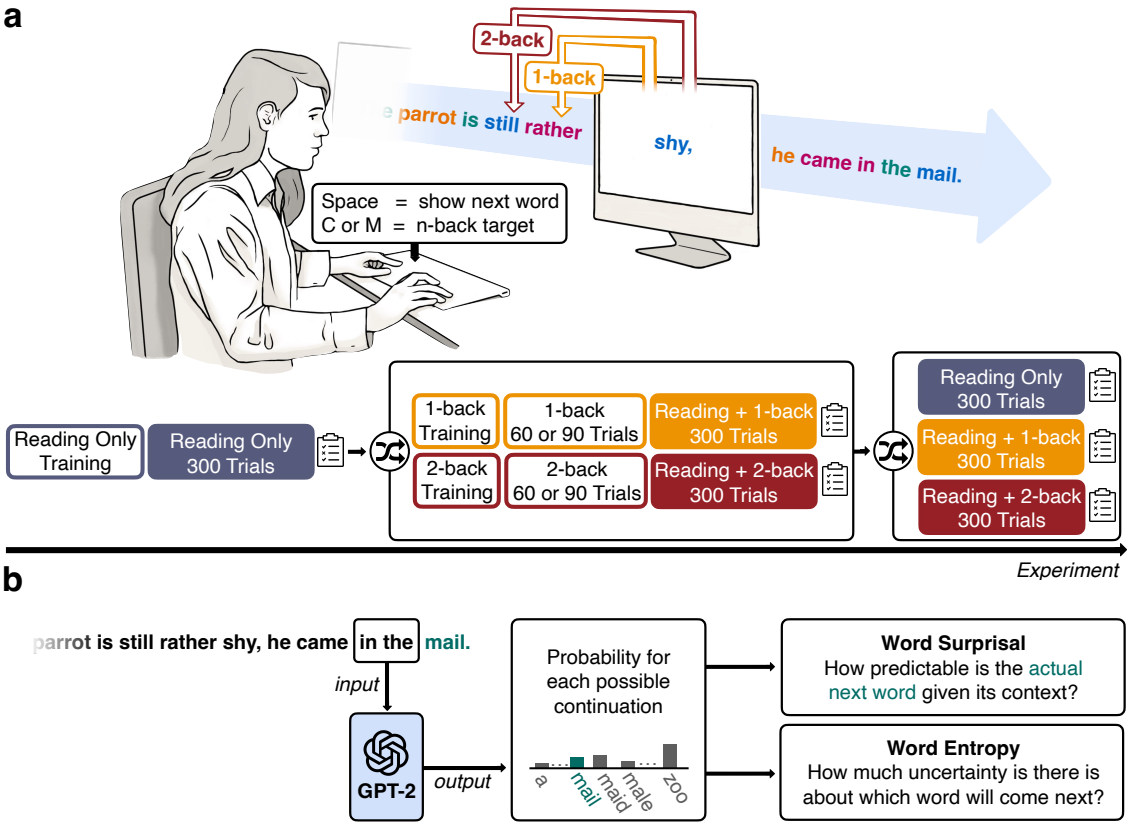

Figure 2. Experimental design and quantification of predictability as word surprisal using GPT-2. **a**. Participants were asked to perform a self-paced reading task (Reading Only) which was complemented in some blocks by a secondary n-back task on the font colour of the words (Reading + 1-back, and Reading + 2-back). The order of the blocks was pseudo-randomized, with Reading Only always being the first condition to be presented, followed by the two dual-task conditions, and another main block for each of the three conditions. Both dual-task paradigms (Reading + 1-back and Reading + 2-back) were first introduced in short single-task training sets. **b.** We generated one surprisal score for each word in the reading material by using context chunks of 2 words as prompts for next-word-predictions in GPT-2. The resulting probability for the actual next word in the text (here: "mail", marked in teal) was then transformed into a surprisal score, which reflected how predictable the respective word was given the context. Additionally, based on the distribution of probabilities for all possible continuations, we computed an entropy score, which reflects the uncertainty in predicting the next word. Please note that the example sentence used here has been translated to English for better comprehensibility, while the original text materials were in German.

Increased Cognitive Load and Older Age Reduce Task Performance

Results from a generalised mixed model (GAM) for accuracy data showed that participants read the texts carefully and answered most comprehension questions correctly, with accuracies of 93% ± 14% (mean ± SD) in the Reading Only, 80% ± 25% in the 1-back, and 73% ± 28% in the 2-back Dual Task conditions (Fig. S2b). Increasing cognitive load (χ^2 = 156.4, p < 0.001) and age (χ^2 = 7.6, p = 0.006) were associated with poorer performance (see *Methods* for model details; Fig. 3a, Table S1). Similarly, d-prime values demonstrated good performance in the n-back task, with mean d-primes of 3.77 ± 0.8 in the 1-back and 2.12 ± 0.87 in the 2-back Dual Task conditions (Fig. S2a). The overall high magnitude of the d-primes observed here can be attributed to the low target ratio in the experiment, resulting in high correct rejection and low false alarm rates. A linear mixed-effects model revealed performance declines with cognitive load (χ^2 = 651.5, p < 0.001) and age (χ^2 = 15.8, p < 0.001; Fig. 3b, Table S1).

Reading Time Increases with Age, Surprisal, and Cognitive Load

To operationalise language predictability, we calculated word surprisal scores using a 12-layered generative pre-trained transformer model (GPT-2; Radford et al., 2019; Fig. 2b). Word surprisal quantifies the predictability of the current word given its preceding context (Hale, 2001; Levy, 2008; Smith and Levy, 2013). We chose a context length of two words, as constraining the context has been shown to increase GPT-2's psychometric predictive power, making its next-word predictions more human-like (Kuribayashi et al., 2022). In addition to each word's surprisal, we also computed word entropy, which reflects the uncertainty in predicting the next word. Thus, while word surprisal indicates the predictability of each word, word entropy reflects the uncertainty underlying its prediction.

We used linear mixed-effects regression to assess the effect of word surprisal on reading times and its interaction with cognitive load and age (see *Methods* for model details). As hypothesised, we observed significantly longer reading times with advancing age, independent of cognitive load condition (β = 0.009, t = 9.199, p < 0.001). For instance, each additional year of age increased reading times by roughly 1% or 9ms, respectively (Fig. 3c, first panel, and Table 1).

Likewise, increased cognitive load was associated with significantly longer reading times ($\beta_{1\text{-}back}$ = 0.474, $t_{1\text{-}back}$ = 34.046, $p_{1\text{-}back} < 0.001$, $\beta_{2\text{-}back}$ = 0.792, $t_{2\text{-}back}$ = 30.339, $p_{2\text{-}back} < 0.001$; Table 1 and Fig. 3c, second panel), indicating participants read more slowly when faced with a more challenging task. Even after excluding the *Reading Only* condition from the model and comparing only the two equally attention-demanding dual-task conditions *1-back*

and *2-back*, hereby controlling for attentional switching costs, this effect still holds true (β = 0.339, t = 16.221, p < 0.001; Table S3).

Moreover, in line with our hypothesis, we observed a consistent increase in reading times with higher surprisal (β = 0.002, t = 11.321, p < 0.001; Table 1 and Fig. 3c, third panel), suggesting highly predictable words were read more quickly than unpredictable ones.

Cognitive Load Reduces the Impact of Surprisal

In line with our hypotheses, cognitive load significantly modulated the effect of surprisal ($\beta_{1\text{-}back}$ = -0.001, $t_{1\text{-}back}$ = -6.772, $p_{1\text{-back}}$ < 0.001, $\beta_{2\text{-}back}$ = -0.001, $t_{2\text{-}back}$ = -7.681, $p_{2\text{-back}}$ < 0.001). While we observed a clear increase in reading time in the *Reading Only* condition when surprisal was high, this effect was mitigated in both the *1-back* and *2-back Dual-Task* condition, where cognitive load was increased (Table 1 and Fig. 3d, first panel).

Age Modulates the Effect of Cognitive Load

While increased task demands were associated with prolonged reading times across participants, we found this effect became more pronounced with advancing age ($\beta_{1\text{-}back}$ = -0.003, $t_{1\text{-}back}$ = -3.606, $p_{1\text{-back}}$ < 0.001, $\beta_{2\text{-}back}$ = -0.002, $t_{2\text{-}back}$ = -1.690, $p_{2\text{-back}}$ = 0.097; Table 1 and Fig. 3d, second panel).

Age Modulates the Effect of Surprisal

Contrary to our hypotheses, older age was associated with stronger surprisal effects (β = -0.00004, t = 9.287, p <0.001; Table 1 and Fig. 3d, third panel). This suggests that, as individuals age, the effect of word predictability on their reading time becomes increasingly pronounced instead of remaining constant or diminishing, as we had originally anticipated. A post-hoc simple slopes analysis resolving this two-way interaction demonstrated that the effect of age on the slope of surprisal was significant across the whole age range of our sample and showed a continuous incline ($\beta_{85\text{ - }18\text{ years}} = 0.002,\ p < 0.001$).

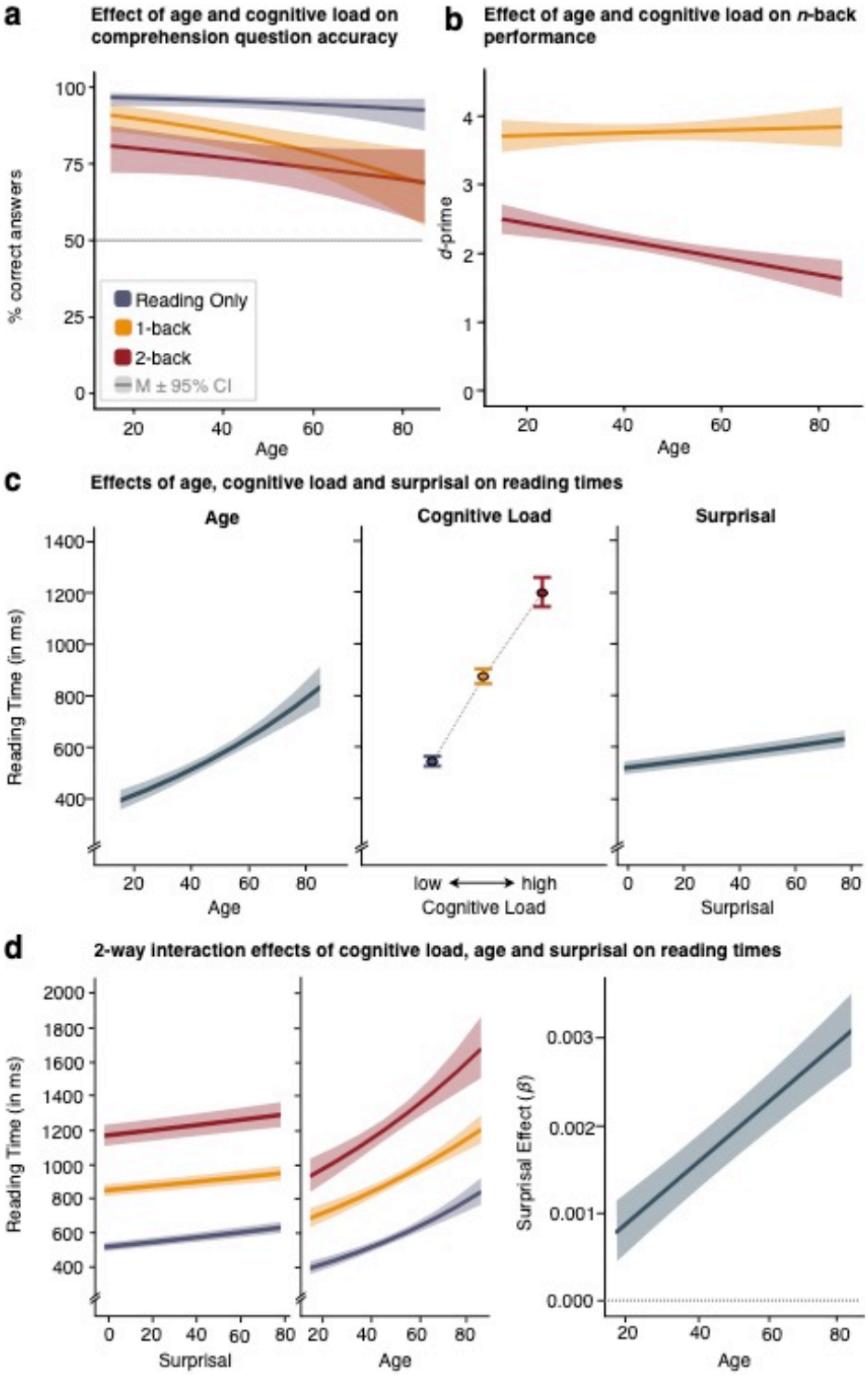

Figure 3. Estimated marginal effects of predictors age, cognitive load and surprisal on task performance and reading time, adjusted for "Reading Only" as the reference level. Main effects of cognitive load and age on accuracy in the comprehension question task (panel **a**) and on n-back task performance (d-primes; panel **b**). Please note that we do not show *d*-primes for the Reading Only task as there was no n-back task in this condition. Reading time increased with increasing age, cognitive load and word surprisal (panel **c**), with cognitive load modulating the effect of surprisal and age (panel **d**, left and middle), and age significantly modulating the effect of surprisal across the full age range (Johnson-Neyman plot in panel **d**, right). For raw and predicted individual trajectories, please see Fig. S2 and S3 in the Supplementary Material.

Age modulates the Interaction of Surprisal and Cognitive Load

Finally, we tested whether the observed interaction between cognitive load and surprisal changes with advancing age: We found that age indeed modulated the joint influence of cognitive load and surprisal on reading times, as reflected by a significant three-way interaction of surprisal, cognitive load and age ($\beta_{1\text{-back - Reading Only}} = -0.00011$, $t_{1\text{-back}} = -12.2661$, $p_{1\text{-back}} < 0.001$, $\beta_{2\text{-back - Reading Only}} = -0.00008$, $t_{2\text{-back}} = -8.484$, $p_{2\text{-back}} < 0.001$; Table 1). Even after excluding the *Reading Only* condition from the model and contrasting only the dual-task conditions *1-back* and *2-back* (Control analysis), results still showed a significant three-way interaction of age, cognitive load and surprisal ($\beta = 0.00003$, $t = 3.373$, $p = 0.001$; Table S3).

To get a more nuanced understanding of this effect, we conducted a simple slopes analysis for our original model (Fig. 4): Under low cognitive load (condition *Reading Only*), surprisal significantly influenced reading times in all but the youngest participants. Put simply, when participants read a text without having to perform an additional n-back task, predictable (i.e., low surprisal) words yielded significantly shorter reading times than unpredictable (i.e., high surprisal) words (Fig. 4, left panel, first plot). This effect was most pronounced in the oldest participants ($\beta_{85\text{ - 18 years}} = 0.006,\ p < 0.001$).

As task demands increased, we observed a reversal of this pattern, with younger participants exhibiting stronger surprisal effects than older participants (condition *1-back*; Fig. 4, left panel, second plot). Notably, under increased cognitive load, even the youngest participants started showing significant surprisal effects ($\beta_{85\text{ - 18 years}} = -0.0009,\ p = 0.04$). Finally, in the most demanding condition (*2-back*), the pattern shifted again, with older adults again showing stronger surprisal effects than their younger counterparts ($\beta_{85\text{ - 18 years}} = 0.001,\ p = 0.003$; Fig. 4, left panel, third plot).

Comparing surprisal effects between cognitive load conditions (e.g., compare conditions *Reading Only* and *1-back* in Fig. 4, left panel, first and second plot) revealed that older adults showed the most pronounced reduction in surprisal effects as cognitive load increased, which suggests they were more vulnerable to increased task demands than younger participants (18 year-old: $\beta_{1\text{-back - Reading Only}} = -0.003,\ p < 0.001,\ \beta_{2\text{-back - Reading Only}} = -0.003,\ p < 0.001,\ \beta_{2\text{-back - 1-back}} = 0.0004,\ p = 0.06$; 85 year-old: $\beta_{1\text{-back - Reading Only}} = -0.011,\ p < 0.001,\ \beta_{2\text{-back - Reading Only}} = -0.008,\ p < 0.001,\ \beta_{2\text{-back - 1-back}} = 0.003,\ p < 0.001$).

Modelling potential non-linear contributions

To account for the possibility that age, surprisal, and their interaction with cognitive load might demonstrate a non-linear effect on reading time, we fitted a generalised additive mixed-effects model (GAM) to our data, using the same model structure as for the linear regression model (see *Methods* for model details). Results from the GAM revealed overall similar effects relative to the LMM. The effective degrees of freedom (EDF) of all continuous predictors were estimated above 1, confirming their non-linearity. Similar to the LMM, both our predictors of interest, surprisal and age, demonstrated significant effects on reading times (Fig. 5), with surprisal showing a more non-linear trajectory than age (EDF for surprisal: 4.107, EDF for age: 3.028, both $p < 0.001$). The smoothing splines for the three-way interaction of age, surprisal, and cognitive load showed significant effects for all levels of cognitive load. Similar to the LMM, we found the strongest effect for the *Reading Only* condition (EDF for Reading Only: 10.248, $p < 0.001$, EDF for 1-back: 2.017, $p = 0.036$, EDF for 2-back: 4.89, $p = 0.024$). The full model results are reported in Supplementary Table S4 and Figure S4.

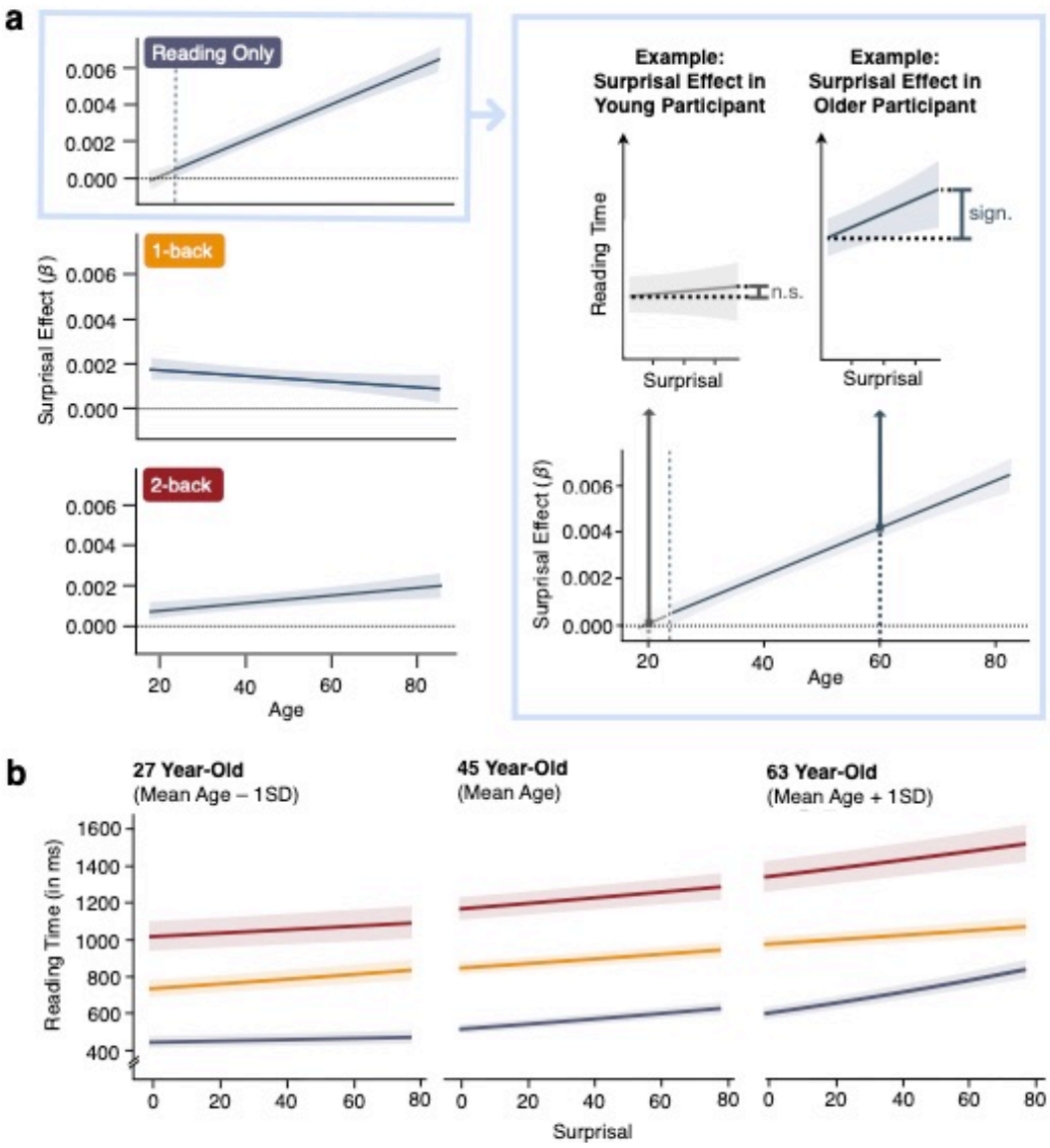

Figure 4. Results of the simple slopes analysis (panel **a**) and exemplary marginal effects plots for three different ages (panel **b**). In the simple slopes plot on the left side of panel **a**, we show the effect of surprisal on reading time across the whole age range separated by cognitive load condition: *Reading Only* (top), *1-back Dual Task* (middle) and *2-back Dual Task* (bottom). The stronger the surprisal effect for a certain age, the higher the value on the y-axis. Grey areas indicate age ranges for which we did not find an effect of surprisal on reading time in the respective condition, whereas blue areas indicate a significant surprisal effect (see inset on the right for a visualisation of a non-significant effect in a younger participant and a significant effect in an older participant). In panel b, we show the predicted surprisal effect in each cognitive load for a young (average age - 1SD), a middle-aged (average age) and an older participant (average age + 1SD).

Table 1

Main results for model for reading times (N = 175)

	Predictors	*Estimates*	*Std. Error*	*CI*	*t*	*p*	
Main Effects	surprisal	0.001707	0.000151	0.001411 – 0.002002	11.320677	**1.368×10^{-28}**	*
	age	0.009113	0.000991	0.007158 – 0.011068	9.199100	**1.751×10^{-16}**	*
	cognitive load [1-back vs. Reading Only]	0.473800	0.013916	0.446336 – 0.501264	34.046321	**8.399×10^{-79}**	*
	cognitive load [2-back vs. Reading Only]	0.791540	0.026090	0.740046 – 0.843034	30.338989	**7.320×10^{-71}**	*
2-way Interactions	surprisal x age	0.000035	0.000004	0.000027 – 0.000042	9.287151	**3.481×10^{-20}**	*
	surprisal x cognitive load [1-back vs. Reading Only]	-0.001093	0.000161	-0.001409 – -0.000776	-6.771521	**2.043×10^{-11}**	*
	surprisal x cognitive load [2-back vs. Reading Only]	-0.001255	0.000163	-0.001575 – -0.000935	-7.681261	**2.709×10^{-14}**	*
	age x cognitive load [1-back vs. Reading Only]	-0.002798	0.000776	-0.004330 – -0.001267	-3.606479	**5.135×10^{-4}**	*
	age x cognitive load [2-back vs. Reading Only]	-0.002458	0.001454	-0.005329 – 0.000412	-1.690400	9.681×10^{-2}	
3-way Interactions	surprisal x age x cognitive load [1-back vs. Reading Only]	-0.000111	0.000009	-0.000129 – -0.000094	-12.266076	**3.748×10^{-34}**	*
	surprisal x age x cognitive load [2-back vs. Reading Only]	-0.000078	0.000009	-0.000096 – -0.000060	-8.483676	**4.384×10^{-17}**	*
Model Fit	Intra-class Correlation (ICC)			0.46			
	Marginal R^2 / Conditional R^2			0.643 / 0.807			

Note. All continuous predictors were centered. Degrees of freedom for *p*-values, standard errors and confidence intervals (CI) were computed using Satterthwaite's approximation. All *p*-values reported here are FDR-corrected and were computed using ANOVAs with type III sum of squares. Results that are significant on an alpha-level of 0.05 are marked with a star.

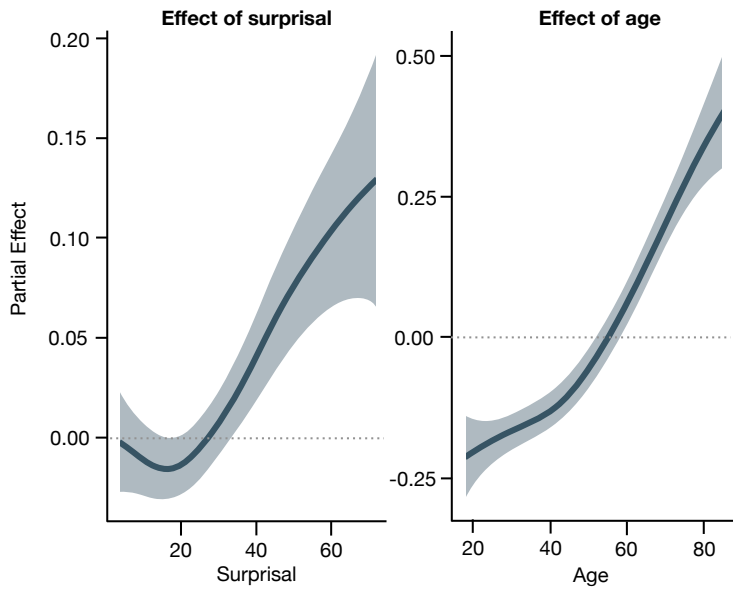

Figure 5. Results of the GAM control analysis for the effects of surprisal and age on reading times. For an illustration of the 3-way interaction of age, cognitive load and surprisal, please see Fig. S4.

Interaction of age, surprisal, and cognitive load generalises to new sample

To probe the reliability of our findings, we conducted an exact online replication of the original experiment. When comparing the results of the online participants of the original sample and the online replication sample, we find similar patterns, as illustrated in Figure 6. We again found a significant 3-way interaction of age, cognitive load and surprisal ($\beta_{1\text{-}back}$ = -0.00002, $t_{1\text{-}back}$ = -1.605, $p_{1\text{-back}}$ = 0.1386, $\beta_{2\text{-}back}$ = -0.00006, $t_{2\text{-}back}$ = -3.871, $p_{2\text{-back}}$ < 0.001; see Table S2), albeit using a smaller sample size of 96 instead of 175 participants and a slightly different age distribution. Notably, the three-way interaction showed similar effect sizes and directions of effects in the replication sample relative to our original online sample.

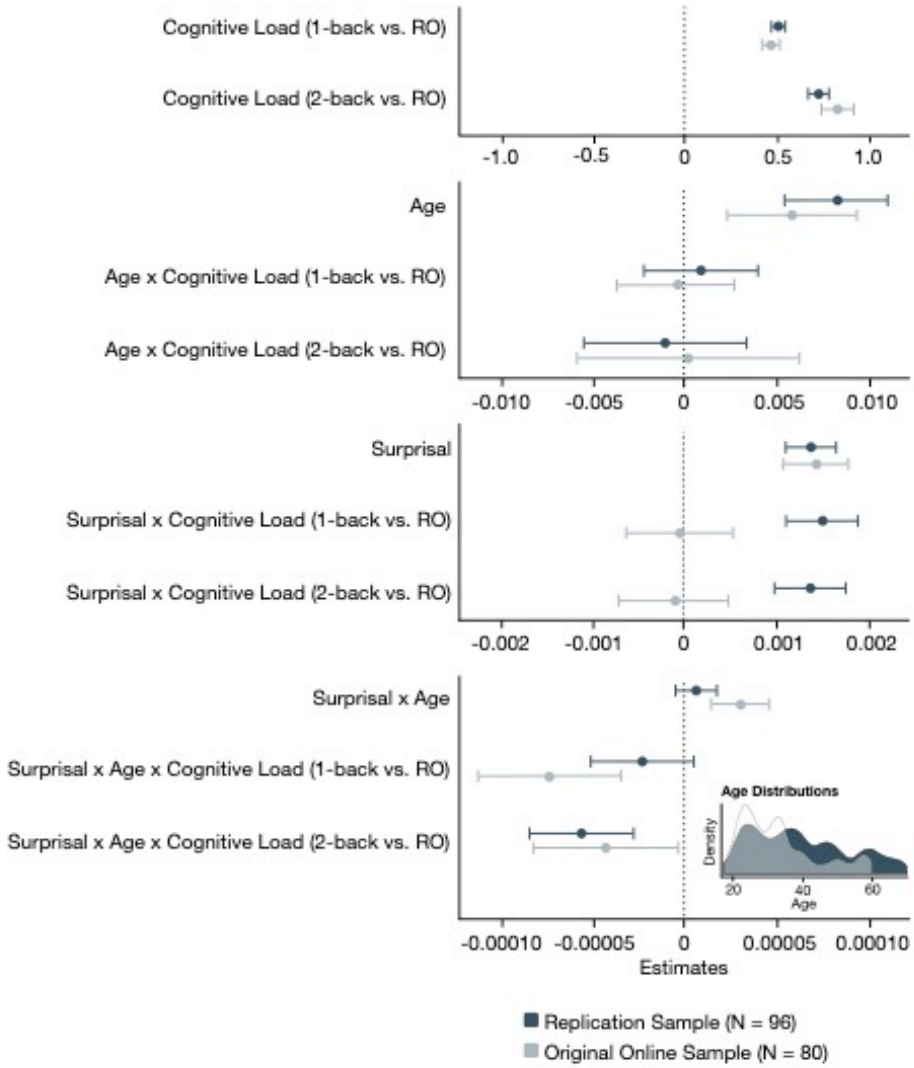

Figure 6. Results of the internal online replication in comparison with the results of the online sample of the original study. Shown are estimates ± CI for all main effects of surprisal, age, and cognitive load as well as their 2-way and 3-way interactions. RO: Reading only. For full results please see Table S2. For a comparison of age distributions in the original online and lab sample and the online replication sample, please see Fig. S1. Please note that effects are grouped by their magnitude.

DISCUSSION

Linguistic predictions are a powerful feature of language comprehension. But do they really come 'for free', or how much do they draw on executive resources? With the present study, we asked how language predictions change with increasing cognitive load, and how this interaction is modulated by age. To do so, we paired a self-paced reading task with a secondary n-back task on the font colour of the words.

First, as expected, and validating our overall approach, high cognitive load was associated with an increase in reading times as well as a decrease in task performance across age groups. This is consistent with previous studies using n-back tasks (Lamichhane et al., 2020; Kwong See & Ryan, 1995).

Next, as hypothesised, higher word surprisal slowed down reading, even when controlling for word length and frequency as well as prediction entropy. This effect of surprisal replicates findings from previous studies showing that highly unpredictable words are generally associated with longer reading times (Heilbron et al., 2022; Monsalve et al., 2012; Schmitt et al., 2021; Shain et al., 2024; Smith & Levy, 2013; Wilcox et al., 2023).

Finally, we explored the relationship between surprisal and cognitive load across the (cross-sectional) adult lifespan. We hypothesised that increasing cognitive load should gradually impair language prediction, which should be reflected by a diminished surprisal effect in conditions with high cognitive load. In line with our hypothesis, we found that the effect of word surprisal on reading times was modulated by cognitive load. Specifically, when cognitive load was high, the effect of surprisal on reading times was significantly diminished.

Interestingly, this interaction between surprisal and cognitive load was significantly modulated by age. While age generally increased the reliance on language prediction, it also increased the susceptibility of this strategy to changes in available executive resources: Older adults showed the strongest relative reduction of the surprisal effect with increasing cognitive load. However, under high load, they still showed the strongest surprisal effect in absolute terms. In a direct replication of the original experiment, we reproduced this finding, further confirming the reliability of our results.

Disentangling effects of attention versus executive resource

When investigating the interaction of cognitive load and language predictability on reading time, we found that increased cognitive load diminished the effect of word predictability, suggesting executive resources are involved in the generation of language predictions. In contrast to previous work on the influence of cognitive load on language comprehension (e.g. Fedorenko et al., 2006; Gordon et al., 2002; Rodd et al., 2010), we employed a *non-linguistic*

secondary task to manipulate cognitive load, thereby aiming to engage domain-general executive resources.

Predictive processing has not only been suggested to be foundational to language comprehension (e.g. Heilbron et al., 2022; Schmitt et al., 2021; Ryskin & Nieuwland, 2023), but it is also thought to be a core mechanism of the human brain (Bubic et al., 2010; Friston, 2010; Rao & Ballard, 1999). Drawing parallels between domains can thus offer valuable insights into the mechanisms of language prediction: For instance, there is evidence from the visual domain showing that attending to the stimulus material is a prerequisite for predictive processing (Larsson & Smith, 2012; Richter & de Lange, 2019). This observation from the visual domain can potentially be extended to our findings, wherein language predictions were diminished if attention had to be divided between the reading material and a challenging non-linguistic secondary task.

However, attention alone cannot fully account for the differences in sensitivity to word predictability shown here, as we observed the predictability effect did not only diminish if a secondary task was introduced, but also decreased with increasing cognitive load even when attentional switching costs were held constant between conditions. In line with previous literature (e.g., Huettig & Janse, 2016; Ito, Corley & Pickering, 2017; Fricke & Zirnstein, 2022), our results thus suggest that executive processes beyond attention, such as updating and maintenance of context information, shifting between tasks, and inhibition of irrelevant information (Friedman & Miyake, 2017), are integral to language prediction.

Language prediction as a compensatory mechanism in older age

Further examining the effect of language predictability across the adult lifespan revealed interesting differences, namely older adults showing stronger predictability effects than younger participants. Our result aligns with previous work on age-related changes in linguistic predictions, indicating heightened sensitivity to unexpected lexical input in older adults (Cheimariou et al., 2019). This finding may reflect a commonly reported pattern of greater reliance on intact vocabulary and world knowledge with age in the face of declining executive functions (Martin et al., 2022; Salthouse et al., 2003). Here, we show that even under increased cognitive load, older adults still rely heavily on their refined system of linguistic predictions driven by their lifelong experience. This allows them to make more fine-tuned predictions, but also renders them more vulnerable to unexpected information than their younger counterparts.

Previous studies have found a larger vocabulary size associated with more rapid processing of language and improved language comprehension (Borovsky et al., 2012;

Huettig & Pickering, 2019; Mainz et al., 2017; Matthews, 2018; Stæhr, 2008 & 2009). In line with this, older adults exhibit more advanced language processing abilities compared to children or younger adults due to their accumulated years of exposure to language and their increased vocabulary (Brysbaert et al., 2016; Ito & Sakai, 2021). This accumulated skill is thought to serve as a compensatory mechanism for decline in working memory capacities or reduced executive functioning with age (Bunzeck, et al., 2024; Reuter-Lorenz et al., 2021). Indeed, speech comprehension appears to remain largely intact in older adults (Burke, & Mackay, 1997; Shafto & Tyler, 2014).

In addition to differences in executive resources, it might be worth considering inter-individual differences regarding the *utilisation* of formed predictions between younger and older adults: While we assumed that formed predictions are utilised automatically, and that the observed differences in the effect of language predictability on reading times between individuals might thus be attributable to a difference in executive resources, one could also argue that individuals might simply weigh their formed predictions differently.

Akin to the longer exposure to language in older individuals, it is reasonable to assume that older individuals have also had more time to accumulate experience regarding the *accuracy* of their predictions, and to refine their predictions through prediction errors. Consequently, older adults might rely more heavily on their predictions than younger adults (Chan et al., 2021; Moran et al., 2014). Additionally, due to age-related sensory decline (Verhaeghen et al., 2003), older adults might exhibit a stronger dependence on context-based predictions to process language, as incoming sensory information might be less informative (Rogers, 2017; Wingfield et al., 2015; Wolpe et al., 2016). A stronger reliance on language predictions could thus serve as a compensatory mechanism to facilitate language comprehension despite sensory decline in older adults.

How can limited executive resources affect language prediction?

The relationship between word predictability and age becomes more nuanced when the impact of cognitive load is also considered: While under low load, we observe the effect of word predictability on reading time increases with advancing age – with the youngest participants showing no predictability effects at all –, this pattern reverses under intermediate load, with younger rather than older participants exhibiting the strongest surprisal effects.

Although this pattern may seem counterintuitive, we interpret it as follows: A non-existent predictability effect in very young adults on the behavioural level does not necessarily indicate an absence of predictive processing if cognitive load is minimal. Rather, a more plausible explanation is that due to their intact executive resources, young adults are so adept

at processing incoming stimuli that even highly unpredictable words can be handled swiftly. This results in minimal differences in reading time between low- and high-surprisal words, eliminating any significant surprisal effects. However, as cognitive load increases and executive resources are diverted, the youngest cohort starts showing predictability effects. With further increases in load, these effects diminish again, suggesting predictive processing is impaired as executive resources become more limited.

The formation of language predictions in young adults under increased task demands necessitates further investigation at the neural level: For instance, observing diminished neural markers of prediction formation – such as pre-stimulus alpha or beta changes in EEG or MEG (Bastos et al., 2012; Gastaldon et al., 2020; Rommers et al., 2017; Sedley et al., 2016; Wang et al., 2018; Terporten et al., 2019) – in conditions with low compared to increased task load would suggest that younger individuals do not form language predictions. For older adults, a different pattern emerges: Under intermediate load, they exhibit a clear reduction in the effect of word predictability on reading time, unlike the increased predictability effect seen in young adults. According to the CRUNCH theory (Reuter-Lorenz & Cappell, 2008), this occurs because older adults reach their "crunch point" – the threshold at which cognitive resources are exhausted and task performance declines – earlier than younger adults, in the current experiment specifically in the intermediate cognitive load condition (*1-back*).

When cognitive load is moderate, older adults still manage to process both the reading task and the n-back task simultaneously, albeit with the n-back task diverting executive resources that would otherwise support predictive processing. When cognitive load increases further, older adults may no longer sustain both tasks simultaneously and might begin alternating between reading and n-back task, effectively treating them as separate single tasks performed in short bursts of only a few trials. This might result in fluctuating behaviour in the reading part of the tasks: In some trials, participants might show predictability effects similar to those seen in the low-load baseline condition *Reading Only*, and other times showing none. An analogous situation possibly arises in the n-back part of the task, with participants achieving performance levels as in a *2-back Single-Task* block in some trials, and near-zero performance when they effectively ignored the n-back task. Alternatively, some older adults may handle both tasks sequentially on a trial-by-trial basis, resulting in extended reading times. Overall, these age-related changes lead to reduced word predictability effects and lower, though still above-chance, n-back performance. Consequently, even after accounting for performance variability and excluding blocks with inadequate performance, older participants still exhibit stronger surprisal effects in the more demanding 2-back condition compared to the less challenging *1-back* condition.

Considering these differences in “crunch points” between younger and older adults, the lowest predictability effect observed in either of the dual-task conditions should be interpreted as a result of depleted executive resources.

CONCLUSION

In summary, the present study helps resolve the debate about the cognitive cost of predictive language processing. The data offers the following key insights:

First, low language predictability as well as high task demands both have a detrimental effect on reading time. This holds true across a large age range. Second, we find that higher task demands diminish the effects of language predictability on reading time, replicably demonstrating that language prediction draws on resources of executive control. Third, the data reveal age-related differences in the use of linguistic predictions: with High predictability had more leverage on reading efficiency in older than in younger adults but was also more sensitive to available executive resources.

MATERIALS & METHODS

Participants

We initially recorded data from 178 participants, who were either tested online (*N* = 83) or in a controlled lab environment (*N* = 95). We excluded data from three participants from the online sample from further analysis, either due to technical issues (*N* = 1) or because they reported having been distracted during the task (*N* = 2). The resulting final sample comprised 175 participants aged 18 to 85 years (M = 44.9 ± 17.9, 18-85 years; original sample) with a balanced gender distribution of 51% female, 47% male, and 2% non-binary identifying participants. All participants were native German speakers with normal or corrected-to-normal vision, and intact colour vision (assessed by a screening test; Ishihara, 1987). Exclusion criteria were dyslexia, illiteracy, a history of psychiatric or neurological disorders or drug abuse. Individuals who had consumed drugs or alcohol immediately prior to the study were not eligible for participation.

All participants from the online sample enlisted through the recruitment platform Prolific, whereas lab participants were recruited via an existing database of the Max-Planck-Institute. Lab participants above the age of 40 performed the Mini-Mental State Examination (MMSE; Folstein et al., 1975) to screen for cognitive impairments. Middle-aged participants (40-59 years) had a mean score of 29.57 ± 0.84, whereas older participants (60-85 years) scored 28.23 ± 1.62 points.

Participants provided their informed consent prior to participation, and received financial compensation (12€/h). The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics committees of the University of Lübeck and Leipzig University, respectively.

Study Design

During the experiment, participants were asked to read short newspaper articles which were adapted from articles from the news archive of a well-known German magazine ("*Der Spiegel*"). The nine selected articles were edited to be easy to understand, neutral in tone and non-emotional in content to avoid any influences of text content on reading time. Additionally, all texts were limited to a length of 300 words (trials). Participants were instructed to read the texts carefully, as content-related multiple-choice questions were asked after each one. The comprehension questions served both as a measure of reading comprehension as well as a motivator to pay close attention to the reading material.

During this self-paced reading task (*Reading Only* task), each word was presented individually on screen and participants proceeded to the next word by pressing the space

bar on their keyboard (Fig. 2a). To ensure that each word was displayed at least briefly, the response window started after the word had been shown for a fixed period of 50 ms. Each word was presented centred on screen in one of four font colours (Hex codes: #D292F3 [lilac], #F989A2 [muted pink], #2AB7EF [cerulean blue] and #88BA3F [leaf green]) against a white background.

In four of the six main blocks, the self-paced reading task was complemented by a secondary n-back task (see Fig. 2a), in which participants were instructed to press a target button on the keyboard ("M" for right-handed, "C" for left-handed participants) whenever the font colour of the current word matched that of the previous (*1-back Dual Task*) or the penultimate word (*2-back Dual Task*). Participants were still required to press the spacebar to advance to the next word after pressing the target button. Reaction times were recorded for both kinds of responses. Neither the *Reading Only* task nor the *1-back* and *2-back Dual Task* blocks were speeded, allowing participants to complete the experiment at their own pace.

Before being introduced to this combination of reading and n-back task (see Fig. 2a), participants could familiarise themselves with the n-back paradigm in short, non-linguistic, single-task blocks comprising coloured rectangles as stimuli. These blocks served as an introduction to the nature of an n-back task, and to quantify participants' working memory abilities. For all main blocks with an n-back task, a target ratio of 16.667% was used. The low target-ratio was chosen to prevent an excessive number of n-back reactions during the dual-task blocks.

Taken together, the experiment comprised three cognitive load conditions: A baseline condition, comprising a reading task without additional n-back task (*Reading Only*), a reading task with an additional 1-back task (*1-back Dual Task*), and a reading task with an additional 2-back task (*2-back Dual Task*). Each condition was presented in two blocks, each comprising 300 trials. For each block, one of nine texts was randomly selected, with no text occurring more than once.

Generation of Word Surprisal and Entropy Scores

The predictability of each word was operationalised via word surprisal (Fig. 2b), which reflects the predictability of the current word given its preceding context (Hale, 2001; Levy, 2008; Smith and Levy, 2013). A word's surprisal score is defined as the negative logarithm of the word's probability given its context (Hale, 2001):

$$surprisal_{w_n} = -log(\, P(\, w_n \mid w_1, \ldots, w_{n-1}\,))$$

If a word w_n has a high surprisal score, its occurrence given its preceding context w_1, w_2,…, w_{n-1} has a low probability, rendering it highly unpredictable (i.e. surprising).

In addition to each word's surprisal, we also computed the entropy of the probability distribution for each predicted word given its context (Fig. 2b), which reflects the uncertainty in predicting the next word. The entropy is defined as the negative sum of the product of the probability of each word in the vocabulary and its respective logarithm probability (Shannon, 1948), or – put simply – as the average surprisal of all possible continuations in the vocabulary (Slaats & Martin, 2023):

$$entropy(p) = -\sum_{w_n \in W} P(w_n \mid w_1, \ldots, w_{n-1})\; log\, P(\, w_n \mid w_1, \ldots, w_{n-1}\,)$$

If entropy is low, only one or a few possible words in the vocabulary are assigned high probabilities of being the actual next word, hence indicating low uncertainty about which word will come next. This is usually the case if the previous context is very restricting. Conversely, if a vast amount of words in the vocabulary would be suited as continuations for the given context, and the probability distribution across words is fairly uniform, word entropy – and thus uncertainty about which word will come next – is high. Taken together, word surprisal signifies the predictability of each word whereas word entropy signifies the uncertainty underlying its prediction.

In the current study, we computed surprisal scores as well as one entropy score for each word in the experimental texts (mean surprisal: 18.165 ± 7.523, mean entropy: 4.067 ± 0.932; see further information in the supplementary material). We chose a context length of two words, as constraining the context has been shown to increase GPT-2's psychometric predictive power, making its next-word predictions more human-like (Kuribayashi et al., 2022).

To generate word entropy and surprisal scores, we used a 12-layered GPT-2 model (Radford et al., 2019), which was pre-trained on German texts by the MDZ Digital Library team (dbmdz) at the Bavarian State Library, and the corresponding tokenizer, both available from the Hugging Face model hub (Schweter, 2021). Scores were calculated using Python version 3.10.12 (Van Rossum & Drake, 2009).

Analysis

Preprocessing

To gauge participants' response accuracy in the n-back task, we computed the detection-prime (*d*') index. This measure quantifies the ability to distinguish between target and non-target stimuli, in our case trials (i.e. words in dual- and rectangles in single-task blocks) with colour repetitions and trials where the current colour doesn't match the colour from the *n*th previous trial. A *d*' value of 0 signifies an inability to discriminate between signal and noise stimuli, suggesting that participants indicated they saw a target in either no or all trials. Thus, we excluded all dual-task blocks with *d*-primes of 0 from further analyses, which affected only five participants. In total, we excluded three main blocks in the online sample and two main blocks in the lab sample.

After each text block, participants were asked to answer three multiple-choice questions as a measure of reading comprehension. For each question, we showed four response options, with only one of them being correct. To ensure participants performed the tasks as intended and read the words on screen, we excluded all blocks where none of the questions were answered correctly. In total, we excluded one *1-back* and seven *2-back Dual Task* blocks from the online sample, as well as four *1-back* and eleven *2-back Dual Task* blocks from the lab sample, from datasets of six and fourteen participants, respectively.

Lastly, we preprocessed the reading time data: First, any trial exhibiting a raw reading time exceeding 5000 ms was considered an extended break and subsequently excluded from further analyses. To further remove outliers, we followed the procedure recommended by Berger and Kiefer (2021) to ensure comparable exclusion criteria for long and short outliers in typically skewed reading times: First, reading times were transformed on block-level using the *Proportion of Maximum Scaling* method (POMS; Little, 2013). The square root of each value was then taken to ensure a symmetric distribution. We transformed the data on block level to account for potential differences in reading time distributions between blocks. Following this, we *z*-transformed the data and excluded all trials from further analysis where *z*-scores fell outside a range of -2 to 2 (Berger & Kiefer, 2021; Cousineau & Chartier, 2010).

Taken together, we excluded 12968 trials (4.117% of all trials from the main blocks) with an average of 74.103 ± 12.654 excluded trials per participant.

To facilitate interpretability of units in the results, we subsequently continued working with the raw reading times, which had been cleaned of outliers at this stage, and log-transformed them for statistical analysis.

Statistical Analysis of n-Back Responses and Comprehension Questions

To ensure the validity of our cognitive load manipulation in the dual-task blocks, we examined whether increased cognitive load induced a decline both in n-back task performance – as indicated by reduced *d*-primes – and the accuracy in the comprehension question task, as reflected by a lower number of correct answers. We employed a linear mixed-effects model (LMM) for d-primes and a logistic linear mixed-effects model (GLMM, logit link function) for comprehension question accuracy.

In both models, we included recording location (online or lab), cognitive load and continuously measured age (centred) in both models as well as the interaction of age and cognitive load as fixed effects. In the model for the *d*-primes, we additionally included measures of comprehension question accuracy (on participant and block level) as well as the block number as fixed effects to control for different response strategies and tiredness effects, respectively. Moreover, we included the mean d-primes from the *1-back* and *2-back Single Tasks* as a working memory measure.

We assigned simple coding schemes to the factors recording location and cognitive load. While the model for d-primes included by-participant random slopes for cognitive load, the model for comprehension question accuracy comprised random intercepts for participants. Both models included random intercepts for participants and texts.

d-prime ~ mean d-prime from single tasks +
mean comprehension question accuracy +
block-level deviation from mean comprehension question accuracy +
recording location + block number +
*age * cognitive load +*
(1 + cognitive load | ID) +
(1 | text number)

Note. Structure of the model for *d*-primes in the n-back task in dual-task blocks. The variable age was centred and the variable cognitive load encompassed only two levels (1-back and 2-back) as there was no n-back task in the Reading Only condition.

comprehension question accuracy ~ recording location +
*age * cognitive load +*
(1 | ID) + (1 | text number)

Note. Structure of the model for accuracy in the comprehension question task. The variable age was centred. We used a binomial family distribution with a logit link function for modelling the comprehension question accuracies.

Statistical Analysis of Reading Times

We explored the effects of cognitive load, age and surprisal as well as their 2-way and 3-way interactions on log-transformed reading times using a linear mixed model (LMM). The model included an interaction of age, surprisal, and cognitive load motivated by our hypotheses as well as additional fixed effects to control for nuisance effects. The final selection of fixed and random effects structure was based on highest R^2 values.

We included the reading time of the previous word as a fixed effect to control for potential nuisance effects such as post-error-slowing following a missed n-back target in the previous trial, or sequential modulation effects if the previous trial was ended prematurely, leading to an extended reading time carried over to the current trial. As we only modelled reading times from trials where we had surprisal scores, the first two trials of each block were not included in the statistical analyses.

As response strategies may differ between individuals, but also within an individual from block to block, we included two different regressors representing these distinct between- vs. within-participant effects on reading time. Between-participant effects were modelled by the individual mean comprehension performance whereas within-participant effects were modelled by the block-level deviation from this mean (cf. Bell et al., 2018; Tune et al., 2021). We further included block-wise *d*-primes and participant-wise mean single-task *d*-primes as a proxy of each participant's working memory capacity. By incorporating block- and participant level performance measures, which are designed to be sensitive to task difficulty, we accounted for the potential variation in perceived task load between age groups or samples. For instance, the 2-back task might present a greater challenge for an older individual compared to their younger counterpart, therefore rendering the tasks not entirely comparable between age groups if not appropriately controlled for.

The remaining fixed effects entailed the recording location (online or lab), word frequency (as estimated using Python's *wordfreq* package; Speer, et al., 2018), word length, the occurrence of n-back reactions, as well as the position of block and trial in the course of the experiment as main effects. Furthermore, we included entropy as a fixed effect to account for the uncertainty in the prediction of the next word. Surprisal and entropy values were weakly correlated ($r = 0.29$, $p < 0.001$).

Finally, we included the three-way interaction of age (as a continuous predictor), cognitive load (on 3 levels: *Reading Only*, *1-back Dual Task* and *2-back Dual Task*; contrasted using a simple coding scheme with the *Reading Only* condition as the reference level), and surprisal score (continuous predictor). This entails the implicit inclusion of all two-way interactions of age, cognitive load and surprisal, as well as the main effects of those variables.

Random effects included random intercepts for participants, the effect of text, the current word, the colour of the current word, and by-participant random slopes for cognitive load. All continuous predictors were centred.

log(RT) ~ RT of previous word +
block-level d-prime + mean d-prime from single tasks +
mean comprehension question performance +
block-level deviation from mean comprehension question performance +
recording location + entropy +
word frequency + word length (without punctuation) +
n-back reaction + block number + trial number +
surprisal * age * cognitive load +
(1 + cognitive load | ID) + (1 | text number) + (1 | word) + (1 | colour)

Note. Model structure. RT = Reading Time, ID = participant.

To gain a more nuanced understanding of the three-way interaction of age, cognitive load and surprisal, we performed a subsequent simple slopes analysis. This analysis allows exploring the interaction of two continuous predictors, in our case quantifying the slope of the surprisal effect in each of the three cognitive load conditions as a function of age. This way, we determined for which age range and cognitive load condition we observed a significant effect of surprisal on reading times.

For all models, including control analyses and the internal replication, *p*-values were obtained using ANOVAs with type III sums of squares. Degrees of freedom for *p*-values and standard errors were estimated using Satterthwaite's approximation for all LMMs, and Wald's approximation for the GLMM and GAM (Luke, 2017; Satterthwaite, 1946). All effects reported are significant on an alpha level of 0.05 after FDR-correction for multiple comparisons (Benjamini & Hochberg, 1995).

All analyses were carried out in *R* version 4.2.2 (R Core Team, 2023) using the packages *gratia*, *interactions, lmerTest, lme4, mgcv*, *modelbased,* and *sjPlot* (Simpson, 2024; Bates et al., 2015; Kuznetsova et al., 2017; Long, 2019; Wood, 2011; Makowski et al., 2020; Lüdecke, 2023).

Control Analysis: Dissociating Cognitive Control from Attention

To disentangle attentional and cognitive load effects, we modelled reading times using an additional linear mixed model of the same structure as described before, but contrasting only the *2-back Dual Task* condition with the less demanding *1-back Dual Task* condition. The two Dual Task conditions only differ in cognitive load, but not attentional switching costs, which means any effects of cognitive load can be attributed to the cognitive load manipulation, with attentional demands held constant.

Control Analysis: Modelling Non-Linear Effects of Age

We also fitted a generalised additive model (GAM) to our data to allow for non-linear relationships of the predictor variables with the reading times, as it has been shown that reaction times are oftentimes modulated by predictor variables in a non-linear way (Miwa & Baayen, 2021; Wood, 2017). Specifically, we employed a GAM to account for a potential non-linear relationship of age and surprisal with reading times. To this end, all continuous predictors were fitted with thin-plate regression splines and the interaction of surprisal, age, and cognitive load was fitted via a tensor product smooth with individual curves for each level of cognitive load. The number of basis dimensions for each smoothing spline was checked via model diagnostics available in *mgcv* after the first model set-up and appropriately updated to reach a k-index > 1.01 and $p > 0.05$ to avoid oversmoothing. The random effect structure was set up similarly to the linear mixed model. Further details of the GAM set-up are available in the supplementary material.

Internal Replication

To ensure the reliability of our findings, we conducted an internal replication of the previously described experiment. This internal replication was preregistered on OSF (doi: 10.17605/OSF.IO/SU6VX).

Replication Sample

As outlined in the preregistration, we conducted an online study with a sample of 100 participants. We excluded data from four participants from further analysis, either due to technical issues ($N = 2$) or because they reported having been distracted during the task ($N = 2$). The resulting final sample comprised 96 participants aged 18 to 70 years (M = 39.750 ± 13.996 years) with a balanced gender distribution of 51% female, 48% male, and 1% non-binary identifying participants. As in the original experiment, all participants were native German speakers with normal or corrected-to-normal vision and intact colour vision without dyslexia, illiteracy, a history of psychiatric or neurological disorders or drug abuse. Individuals

who had consumed drugs or alcohol immediately prior to the study were not eligible for participation.

All participants were recruited online through Prolific and only tested upon having provided their informed consent. Participants received a financial compensation of 12€/h. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics committee of the University of Lübeck.

Replication Analyses

Analogous to the original experiment, we first cleaned the reading time data of trials exceeding a duration of 5000 ms as well as outliers (see section *Preprocessing*), which affected 4.036% of all trials from the main blocks with an average of 72.656 ± 11.089 excluded trials per participant. The structure of the statistical model for the replication analysis was analogous to the model for the original analysis of reading times (see section *Statistical Analysis of Reading Times*), except for the predictor of recording location, which was excluded as we only analysed data collected online. To compare results from the original experiment and the replication, we fitted the model once using the data of the online sample from the original experiment and once using the new online replication datasets.

Given that we simplified the analysis approach in the original study after having preregistered the replication, we deviated from the analysis plan described in the preregistration and made the same modifications here, resulting in the use of word surprisal for only one context length instead of four and, consequently, only one LMM instead of several.

Acknowledgements

We thank Christian Koblitz and Marcel Blumenthal for their help with data acquisition in the lab.

Funding

This work was supported by German Research Foundation (DFG, OB 352/2-2 to JO, and HA 6314/4-2 to GH). GH was supported by Lise Meitner Excellence funding from the Max Planck Society and the European Research Council (ERC-2021-COG 101043747).

Author contributions

M.S.: Conceptualization, data curation, formal analysis, investigation, methodology, project administration, visualisation, writing: original draft, review & editing

S.M.: Conceptualization, data curation, formal analysis, investigation, methodology, project administration, writing: original draft, review & editing

S.T.: formal analysis, methodology, writing: review & editing

L-M.S.: methodology, writing: review & editing

G.H.: Conceptualization, funding acquisition, supervision, methodology, writing: review & editing

J.O.: Conceptualization, methodology, funding acquisition, supervision, writing: review & editing

Declaration of AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 3.5 in order to rephrase sentences. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Competing interests

The authors declare that they have no competing interests.

Data Availability

All experimental and analysis scripts are publicly available in the project's Github repository https://github.com/MMarieSchuckart/EXNAT-1. Raw data as well as model outputs, results, and model comparisons will be made available upon publication in the project's OSF repository: https://osf.io/dtwjk/.

Preregistration

The internal replication of the main experiment was preregistered on OSF (doi: 10.17605/OSF.IO/SU6VX).

References

Bamberger, R., & Vanecek, E. (1984). Lesen – Verstehen – Lernen – Schreiben: Die Schwierigkeitsstufen von Texten in deutscher Sprache. Jugend u. Volk Sauerländer.

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

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695–711. doi: 10.1016/j.neuron.2012.10.038

Bell, A., Fairbrother, M. & Jones, K. (2018). Fixed and random effects models: making an informed choice. *Quality & Quantity*, *53,* 1051–1074. doi: 10.1007/s11135-018-0802-x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. doi: 10.2307/2346101

Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, *12*, 675558. doi: 10.3389/fpsyg.2021.675558

Borovsky, A., Elman, J. L., Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417–436. doi: 10.1016/j.jecp.2012.01.005

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research, 2*(1). doi: 10.16910/jemr.2.1.1

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology, 7*. doi: 10.3389/fpsyg.2016.01116

Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience, 4*. doi: 10.3389/fnhum.2010.00025

Bunzeck, N., Steiger, T. K., Krämer, U. M., Luedtke, K., Marshall, L., Obleser, J., & Tune, S. (2024). Trajectories and contributing factors of neural compensation in healthy and pathological aging. *Neuroscience and biobehavioral reviews*, *156*, 105489. doi: 10.1016/j.neubiorev.2023.105489

Burke, M. D., & Mackay, D. G. (1997). Memory, language, and ageing. *Philosophical Transactions of the Royal Society B, 352*(1363), 1845–1856. doi: 10.1098/rstb.1997.0170

Chan, J. S., Wibral, M., Stawowsky, C., Brandl, M., Helbling, S., Naumer, M. J., Kaiser, J. & Wollstadt, P. (2021). Predictive coding over the lifespan: Increased reliance on perceptual priors in older adults—A magnetoencephalography and dynamic causal modeling tudy. *Frontiers in Aging Neuroscience*, *13*. doi: 10.3389/fnagi.2021.631599

Cheimariou, S., Farmer, T. A., & Gordon, J. K. (2018). Lexical prediction in the aging brain: The effects of predictiveness and congruency on the N400 ERP component. *Aging, Neuropsychology, and Cognition*, *26*(5), 781–806. doi: 10.1080/13825585.2018.1529733

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204. doi: 10.1017/S0140525X12000477

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal Of Psychological Research*, *3*(1), 58–67. doi: 10.21500/20112084.844

Dave, S., Brothers, T., & Swaab, T. (2018). 1/f neural noise and electrophysiological indices of contextual prediction in aging. *Brain Research*, *1691*, 34–43. doi: 10.1016/j.brainres.2018.04.007

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *33*(26), 10688–10697. doi: 10.1523/JNEUROSCI.4596-12.2013

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language, 54*(4), 541–553. doi: 10.1016/j.jml.2005.12.006

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). *Mini-mental state examination (MMS, MMSE)* [Database record]. APA PsycTests. doi: 10.1037/t07757-000

Fricke, M., & Zirnstein, M. (2022). Predictive processing and inhibitory control drive semantic enhancements for non-dominant language word recognition in noise. *Languages, 7*(3), 239. doi: 10.3390/languages7030239

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. doi: 10.1016/j.cortex.2016.04.023

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience,* 11, 127–138. doi: 10.1038/nrn2787

Gastaldon, S., Arcara, G., Navarette, E., & Peressotti, F. (2020). Commonalities in alpha and beta neural desynchronizations during prediction in language comprehension and production. *Cortex, 133,* 328–345. doi: https://doi.org/10.1016/j.cortex.2020.09.026

Geranmayeh, F., Chau, T. W., Wise, R. J. S., Leech, R., & Hampshire, A. (2017). Domain-general subregions of the medial prefrontal cortex contribute to recovery of language after stroke. *Brain*, *140*(7), 1947–1958. doi: 10.1093/brain/awx134

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430. doi: 10.1111/1467-9280.00475

Hale, J. T. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the North American Chapter of the Association for Computational Linguistics on Language Technologies, 2*, 159 - 166. doi: 10.3115/1073336.1073357

Häuser, K. I., Demberg, V., & Kray, J. (2019). Effects of aging and dual-task demands on the comprehension of less expected sentence continuations: Evidence from pupillometry. *Frontiers in Psychology*, *10*, 709. doi: 10.3389/fpsyg.2019.00709

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

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience, 31*(1), 80–93. doi: 10.1080/23273798.2015.1047459

Huettig, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences, 23*(6), 464–475. doi: 10.1016/j.tics.2019.03.008

Idowu, M. I., & Szameitat, A. J. (2023). Executive function abilities in cognitively healthy young and older adults – A cross-sectional study. *Frontiers in Aging Neuroscience, 15,* 976915. doi: 10.3389/fnagi.2023.976915

Ishihara, S. M. (1987). Ishihara's tests for colour-blindness. Tokyo: Kanehara & Co.

Ito, A., Corley, M., & Pickering, M. J. (2017). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition, 21*(2), 251–264. doi: 10.1017/S1366728917000050

Ito, A., & Sakai, H. (2021). Everyday language exposure shapes prediction of specific words in listening comprehension: A visual world eye-tracking study. *Frontiers in Psychology, 12*. doi: 10.3389/fpsyg.2021.607474

Kuribayashi, T., Oseki, Y., Brassard, A. & Inui, K. (2022). Context limitations make neural language models more human-like. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, United Arab Emirates*, 10421–10436. doi: 10.48550/arxiv.2205.11463

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. doi: 10.18637/jss.v082.i13.

Kwong See, S. T., & Ryan, E. B. (1995). Cognitive mediation of adult age differences in language performance. *Psychology and Aging, 10*(3), 458–468. doi: 10.1037/0882-7974.10.3.458

Lamichhane, B., Westbrook, A., Cole, M. W., & Braver, T. S. (2020). Exploring brain-behavior relationships in the n-back task. *NeuroImage, 212,* 116683. doi: 10.1016/j.neuroimage.2020.116683

Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: Neuronal adaptation or stimulus expectation? *Cerebral Cortex*, 22(3), 567–576. doi: 10.1093/cercor/bhr119

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177. doi: 10.1016/j.cognition.2007.05.006.

Little, T. D. (2013). *Longitudinal Structural Equation Modeling* (Methodology in the Social Sciences). The GuilfordPress.

Liu, X., Chin, J., Payne, B. R., Fu, W. T., Morrow, D. G., & Stine-Morrow, E. A. (2016). Adult age differences in information foraging in an interactive reading environment. *Psychology and aging*, *31*(3), 211–223. https://doi.org/10.1037/pag0000079

Long, J.A. (2019). *interactions: Comprehensive, user-friendly toolkit for probing interactions*. R package version 1.1.0, https://cran.r-project.org/package=interactions

Lüdecke, D. (2023). *sjPlot: Data visualization for statistics in social science*. R package version 2.8.15. Retrieved from https://CRAN.R-project.org/package=sjPlot

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502. doi: 10.3758/s13428-016-0809-y.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive psychology*, *88*, 22–60. doi: 10.1016/j.cogpsych.2016.06.002

MacGregor, L., Gilbert, R. A., Balewski, Z., Mitchell, D. N., Erzinçlioğlu, S., Rodd, J. M., Duncan, J., Fedorenko, E. & Davis, M. H. (2022). Causal contributions of the domain-general (multiple demand) and the language-selective brain networks to perceptual and semantic challenges in speech comprehension. *Neurobiology of Language*, *3*(4), 665–698. doi: 10.1162/nol_a_00081

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology, 8*. doi: 10.3389/fpsyg.2017.01164

Makowski, D., Ben-Shachar, M., Patil, I., & Lüdecke, D. (2020). Estimation of model-based predictions, contrasts and means. *CRAN*. Retrieved from https://github.com/easystats/modelbased

Martin, S., Saur, D., & Hartwigsen, G. (2022). Age-dependent contribution of domain-general networks to semantic cognition. *Cerebral Cortex*, *32*(4), 870–890. Retrieved from https://doi.org/10.1093/cercor/bhab252

Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System, 72,* 23–36. doi: 10.1016/j.system.2017.10.005

Miwa, K., & Baayen, H. (2021). Nonlinearities in bilingual visual word recognition: An introduction to generalized additive modeling. *Bilingualism: Language and Cognition*, *24*(5), 825–832. doi:10.1017/S1366728921000079

Monsalve, I. F., Frank, S. L., & Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, France,* 398–408.

Moran, R. J., Symmonds, M., Dolan, R. J., & Friston, K. J. (2014) The Brain Ages Optimally to Model Its Environment: Evidence from Sensory Learning over the Adult Lifespan. *PLoS Computational Biology, 10*(1), e1003422. doi: 10.1371/journal.pcbi.1003422

Onnis, L., Lim, A., Cheung, S., & Huettig, F. (2022). Is the mind inherently predicting? Exploring forward and backward looking in language processing. *Cognitive Science, 46*(10), e13201. doi: 10.1111/cogs.13201

Payne, B. R., & Federmeier, K. D. (2018). Contextual constraints on lexico-semantic processing in aging: Evidence from single-word event-related brain potentials. *Brain Research, 1687,* 117–128. doi: 10.1016/j.brainres.2018.02.021

Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and hearing*, *39*(2), 204–214. doi: 10.1097/AUD.0000000000000494

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*(3), 105-110. doi: 10.1016/j.tics.2006.12.002

R Core Team (2023). *R: A language and environment for statistical computing. R Foundation for Statistical Computing,* Vienna, Austria. Retrieved from https://www.R-project.org/

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog.* Retrieved from https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf

Rao, R., & Ballard, D. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience,* 2, 79–87. doi: 10.1038/4580

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation Hypothesis. *Current Directions in Psychological Science, 17*(3), 177–182. doi: 10.1111/j.1467-8721.2008.00570.x

Reuter-Lorenz, P. A., Festini, S. B. & Jantz, T. K. (2021). Executive functions and neurocognitive aging. In K. W. Schaie & S. L. Willis (Hrsg.), *Handbook of the Psychology of Aging* (9th Ed., pp. 67–81). Academic Press. doi: 10.1016/B978-0-12-816094-7.00019-2

Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. *eLife, 8*, e47869. doi: 10.7554/eLife.47869

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language, 115*(3), 182–188. doi: 10.1016/j.bandl.2010.07.005

Rogers, C. (2017). Semantic priming, not repetition priming, is to blame for false hearing. *Psychonomic Bulletin and Review, 24*, 1194–1204. doi: 10.3758/s13423-016-1185-4

Roland, D., Yun, H., Koenig, J.-P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition, 122*(3), 267–279. doi: 10.1016/j.cognition.2011.11.011

Rommers, J., Dickson, D. S., Norton, J. J. S., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience, 32*(5), 576-589, doi: 10.1080/23273798.2016.1183799

Rubenstein, H., & Aborn, M. (1958). Learning, prediction, and readability. *Journal of Applied Psychology, 42*(1), 28–32. doi:10.1037/h0039808

Ryskin, R. & Nieuwland, M. S. (2023). Prediction during language comprehension: What is next? *Trends in Cognitive Sciences, 27*(11), 1032–1052. doi: 10.1016/j.tics.2023.08.003

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*(4), 566–594. doi: 10.1037/0096-3445.132.4.566

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110–114. doi: 10.2307/3002019

Schmitt, L., Erb, J., Tune, S., Rysop, A., Hartwigsen, G. & Obleser, J. (2021). Predicting speech from a cortical hierarchy of event-based time scales. *Science Advances, 7*(49). doi: 10.1126/sciadv.abi6070

Schweter, S. (2021). GPT-2 [dbmdz/german-gpt-2]. In *Hugging Face Model Hub*. Retrieved May 16, 2023, from https://huggingface.co/dbmdz/german-gpt2

Sedley, W., Gander, P. E., Kumar, S., Kovach, C. K., Oya, H., Kawasaki, H., Howard, M. A., III., & Griffiths, T. D. (2016). Neural signatures of perceptual inference. *eLife*, *5*, e11476. doi: 10.7554/eLife.11476

Shain, C., Blank, I. A., Fedorenko, E., Gibson, E., & Schuler, W. (2022). Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *Journal of Neuroscience, 42*(39), 7412-7430. doi: 10.1523/JNEUROSCI.1894-21.2022

Shain, C., Meister, C., Pimentel, T. D., Cotterell, R., & Levy, R. (2024). Large-scale evidence for logarithmic effects of word predictability on reading time. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, *121*(10). doi: 10.1073/pnas.2307876121

Shafto, M. A., & Tyler, L. K. (2014). Language in the aging brain: The network dynamics of cognitive decline and preservation. *Science, 346,* 583-587. doi: 10.1126/science.1254404

Shannon, C. E. (1948). A Mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Simpson, G. (2024). *gratia: Graceful ggplot-Based Graphics and Other Functions for GAMs Fitted using mgcv*. R package version 0.9.2, Retrieved from https://gavinsimpson.github.io/gratia/

Slaats, S., & Martin, A. E. (2023). What's surprising about surprisal. *PsyArXiv*. doi: 10.31234/osf.io/7pvau

Sliwinska, M. W., Violante, I. R., Wise, R. J. S., Leech, R., Devlin, J. T., Geranmayeh, F., & Hampshire, A. (2017). Stimulating multiple-demand cortex enhances vocabulary learning. *Journal of Neuroscience*, *37*(32), 7606–7618. doi: 10.1523/JNEUROSCI.3857-16.2017

Smiler, A. P., Gagne, D. D., & Stine-Morrow, E. A. (2003). Aging, memory load, and resource allocation during reading. *Psychology and Aging*, *18*(2), 203–209. doi: 10.1037/0882-7974.18.2.203

Smith, N. J. & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. Retrieved from https://doi.org/10.1016/j.cognition.2013.02.013

Speer, R., Chin, J., Lin, A., Jewett, S., & Nathan, L. (2018, October 3). LuminosoInsight/wordfreq: v2.2. Zenodo. doi: 10.5281/zenodo.1443582

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal, 36*(2), 139-152, doi: 10.1080/09571730802389975

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*(4), 577–607. doi: 10.1017/S0272263109990039

Terporten, R., Schoffelen, J.-M., Dai, B., Hagoort, P., & Kösem, A. (2019). The relation between alpha/beta oscillations and the encoding of sentence induced contextual information. *Scientific Reports, 9*(1), 20255. doi: 10.1038/s41598-019-56600-x

Tune, S., Alavash, M., Fiedler, L., & Obleser, J. *(2021).* Neural attentional-filter mechanisms of listening success in middle-aged and older individuals. *Nature Communications,* 12, 4533. doi: 10.1038/s41467-021-24771-9

Vaden, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R. & Eckert, M. A. (2013). The cingulo-opercular network provides word-recognition benefit. *Journal Of Neuroscience*, *33*(48), 18979–18986. doi: 10.1523/jneurosci.1417-13.2013

Vaden, K. I., Kuchinsky, S. E., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2015). Cortical Activity Predicts Which Older Adults Recognize Speech in Noise and When. *Journal of Neuroscience*, 35(9), 3929–3937. doi: 10.1523/JNEUROSCI.2908-14.2015.

Vaden, K. I., Jr, Kuchinsky, S. E., Ahlstrom, J. B., Teubner-Rhodes, S. E., Dubno, J. R., & Eckert, M. A. (2016). Cingulo-Opercular Function During Word Recognition in Noise for Older Adults with Hearing Loss. *Experimental aging research*, *42*(1), 67–82. doi: 10.1080/0361073X.2016.1108784

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., & Cerella, J. (2003). Aging and dual-task performance: A meta-analysis. *Psychology and Aging, 18*(3), 443–460. doi: 10.1037/0882-7974.18.3.443

Wang, L., Hagoort, P., & Jensen, O. (2018). Language prediction is reflected by coupling between frontal gamma and posterior alpha oscillations. *Journal of Cognitive Neuroscience, 30*(3), 432-447. doi: 10.1162/jocn_a_01190

Wilcox, E. G., Pimentel, T., Meister, C., Cotterell, R., & Levy, R. P. (2023). Testing the predictions of surprisal theory in 11 languages. *Transactions of the Association for Computational Linguistics*, 11, 1451–1470. doi: 10.1162/tacl_a_00612

Wingfield, A., Amichetti, N. M. & Lash, A. (2015). Cognitive aging and hearing acuity: Modeling spoken language comprehension. *Frontiers in Psychology*, *6*. doi: 10.3389/fpsyg.2015.00684

Wingfield, A. & Grossman, M. (2006). Language and the aging brain: Patterns of neural compensation revealed by functional brain imaging. *Journal Of Neurophysiology*, *96*(6), 2830–2839. doi: 10.1152/jn.00628.2006

Wolpe, N., Ingram, J., Tsvetanov, K., Geerlings, L., Kievit, R. A., Henson, R. N., Wolpert, D. M., Cam-CAN, & Rowe, J. B. (2016). Ageing increases reliance on sensorimotor prediction through structural and functional differences in frontostriatal circuits. *Nature Communications, 7,* 13034. doi: 10.1038/ncomms13034

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3-36. doi: 10.1111/j.1467-9868.2010.00749.x

Wood, S.N. (2017). *Generalized Additive Models: An Introduction with R (2nd ed.)*. CRC Press. doi: 10.1201/9781315370279

Xie, Z., Brodbeck, C., & Chandrasekaran, B. (2023). Cortical tracking of continuous speech under bimodal divided attention. *Neurobiology of Language, 4*(2), 318–343. doi: 10.1162/nol_a_00100