AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychology and Aging


ISSN: 0882-7974

2024, Vol. 39, No. 7, 715–730
https://doi.org/10.1037/pag0000845

# Age Differences in Context Use During Reading and Downstream Effects on Recognition Memory

Katja I. Haeuser and Jutta Kray
Department of Psychology, Saarland University
Collaborative Research Center Information Density and Linguistic Encoding (Sonderforschungsbereich 1102), Saarland University

It is well-known that sentential context modulates sentence processing. But does context also have effects that extend beyond the immediate moment, for example, by impacting the memory representations that people store? And are there age-related differences in this process? Here, we investigated this question. German readers who varied in age self-paced through constraining sentences that continued in a predictable or less predictable fashion. Participants' recognition memory was then tested for previously seen (i.e., "old") words and for initially predictable but not actually presented words (i.e., "lures"). The results showed that readers of all ages slowed down when reading unpredictable sentences. However, aging individuals maintained less sentence-specific information than younger adults: They not only understood sentential materials less correctly on the fly, but they also showed disproportionate rates of false remembering and less successful old–new discrimination in the recognition memory test. Of note, rates of false remembering were reduced in those aging readers who allocated more time toward reading unpredictable sentence continuations. Together, our results show that aging increases reliance on gist or schema-congruent processing but that more attentive encoding of text can buffer against some of the resulting memory distortions.

***Public Significance Statement***

We show that aging increases the likelihood of falsely remembering episodes that were never witnessed but match ones knowledge about the world (i.e., schemas). Aging individuals are also less likely to correctly recognize information that does *not* align with their schemas. This shows that older individuals are more likely to rely on things they already know about the world, but they are less likely to spend time with new information. Finally, we found that aging individuals have a distorted perception of their own memories, as they frequently issue false-memory judgments with high confidence.

*Keywords:* prediction, sentence comprehension, reading, false memory, aging

*Supplemental materials:* https://doi.org/10.1037/pag0000845.supp

Sentence and text processing requires that we quickly and continuously decode and map visual word-level features onto an incrementally accumulating representation of meaning (Kintsch et al., 1990). The efficiency of this process likely changes with increasing age, as older adults often show declines in processing speed and working memory capacity (e.g., Salthouse, 1992, 1994; West, 1996) but at the same time have a greater cumulative experience in lexical processing than younger readers due to their life-long experience with language and text (e.g., Payne et al., 2014; Stine-Morrow et al., 2008). A linguistic feature that is known to alleviate some of the burdens of language processing is context. Indeed, words are processed more effortlessly when they are congruent with, and predictable from, prior context (e.g., Altmann & Kamide, 1999; DeLong et al., 2005; Federmeier & Kutas, 1999; Huettig, 2015; Ito et al., 2016; Kuperberg & Jaeger, 2016; Ryskin & Nieuwland, 2023; Van Berkum et al., 2005). But how does

This article was published Online First August 29, 2024.
Elizabeth A. L. Stine-Morrow served as action editor.
Katja I. Haeuser https://orcid.org/0000-0001-6553-3551
All data files and analysis scripts can be found on this article's project page at the Open Science Framework using the link https://osf.io/czvb9/?view_only=28714207ccfa4f5ca2a7d4848ca11a5e. Parts of the results reported in this article were presented at the Annual Meeting of the Cognitive Science Society 2022.
The authors declare no competing interests. This work was funded by the Deutsche Forschungsgemeinschaft, German Research Foundation—Project-ID 232722074–SFB 1102, Project A5.

Katja I. Haeuser conceived of the study. Katja I. Haeuser collected, analyzed, and visualized the data. Katja I. Haeuser wrote the article. Jutta Kray provided feedback on the article.
Correspondence concerning this article should be addressed to Katja I. Haeuser, Department of Psychology, Saarland University, Campus A 1.3, Room 2.13, 66123 Saarbrucken, Germany. Email: khaeuser@coli.uni-saarland.de

contextual knowledge that is built during sentence comprehension affect long-term memory? And how does aging modulate this relationship when we know that the cognitive profiles of older adults show evidence for both decline and reserve? Here, we address these questions by investigating age-related differences in context use during reading and recognition memory. We begin by examining the evidence on sentence comprehension and memory in young- and old-adult readers.

## Sentence Comprehension and Its Effects on Recognition Memory in Younger Adults

Language users often use context to facilitate processing of incoming words in sentence comprehension. Consequently, words that are predictable from context elicit shorter reading times and reduced N400 event-related potential components (e.g., Kutas & Federmeier, 2011; Staub, 2015) compared to words that are unpredictable. Indeed, prediction, that is, the top–down preactivation of linguistic features before these are perceived, has been identified as a hallmark of sentence processing that may facilitate, or even enable, language learning (e.g., Henson & Gagnepain, 2010; Martin et al., 2018; for reviews on prediction, see Huettig, 2015; Kuperberg & Jaeger, 2016; Ryskin & Nieuwland, 2023). But what happens when contextual predictions are not borne out? Do predictable words become suppressed, or do they remain active in memory? Previous studies including young-adult samples have obtained conflicting evidence. One line of research argues that disconfirmed words are maintained in memory. For example, in a seminal study by Hubbard et al. (2019), young-adult participants of English read sentences (e.g., "Be careful, the top of the stove is very greasy") that constrained their expectations toward a specific lure word (e.g., "hot"). In a subsequent recognition memory test, which asked participants to judge words as "old" or "new," participants were more likely to incorrectly endorse contextually predictable lures as old, compared to new words that had not appeared in the sentence. The authors concluded that predictable words "linger" in memory, eliciting false recognition (also see Chung & Federmeier, 2023; Haeuser & Kray 2022a; Höltje & Mecklinger, 2022; Rich & Harris, 2021; Rommers & Federmeier, 2018a). In contrast to this, other findings have suggested that young-adult comprehenders inhibit initial predictions immediately once they are disconfirmed (e.g., Ness & Meltzer-Asscher, 2018). For example, in a cross-modal priming experiment, young-adult participants showed reduced priming effects for predictable words (e.g., "restaurant") when a previously seen sentence was completed by an unpredictable word compared to when it was not completed by any word (e.g., "John aspires to be a chef and soon wants open up his own … bakery"). Thus, previous research is conflicting regarding whether young-adult language users maintain or suppress contextual predictions.

In addition to investigating the effects of sentence comprehension on false memory, previous studies have also sought to uncover how ongoing processes during comprehension can affect memory for sentence content (i.e., "true" recognition or recall memory). One of the central questions in these studies is what promotes human memory—novelty (i.e., prediction error) or congruency with an existing schema (i.e., high predictability; see Van Kesteren et al., 2012). Error-driven learning accounts argue that incongruency drives memory, whereas schema-congruency accounts argue that consistency with accumulated knowledge about the world improves memory. Critically, previous studies on young-adult samples have obtained evidence supporting both views (e.g., Brod et al., 2013; Brod & Shing, 2019; Corley et al., 2007; Federmeier et al., 2007; Haeuser & Kray, 2021; Heikkilä et al., 2015; Höltje et al., 2019; Hubbard et al., 2019; Hunt & Lamb, 2001; Packard et al., 2017; Rommers & Federmeier, 2018b; see also Tibon et al., 2017). Recent research accommodates these findings by postulating two different brain systems that are differently engaged when encoding schema-congruent versus schema-incongruent material (e.g., Greve et al., 2019; Quent et al., 2022). Thus, human memory is likely driven by congruency with schema *and* prediction error.

## Sentence Comprehension and Its Effects on Recognition Memory in Older Adults

Similarly to younger adults, older adults can and do use linguistic context during sentence comprehension to predict upcoming words (for a review, see Payne & Silcox, 2019), but diverging strands of research have obtained conflicting evidence to what extent and how readily older adults do so (e.g., Federmeier et al., 2002; Noh & Stine-Morrow, 2009; Payne & Federmeier, 2018; Pichora-Fuller, 2008; Pichora-Fuller et al., 1995; Rayner et al., 2006; Stine-Morrow et al., 1996, 1999, 2006; Tun & Wingfield, 1994; Wingfield et al., 1985, 1991; Wingfield & Stine-Morrow, 2000; Wlotko et al., 2010, 2012). Relevant to the present study are findings showing that, as long as working-memory demands are not overly taxed (in which case context use goes down; e.g., Dagerman et al., 2006) or stimulus materials are visually or acoustically degraded (e.g., Pichora-Fuller et al., 1995), older adults are, in principle, able to use linguistic context (e.g., Balota & Duchek, 1991; Light et al., 1991, 1992), but it is currently unclear how readily they do so (e.g., Federmeier et al., 2002; Noh & Stine-Morrow, 2009; Payne & Federmeier, 2018; Wlotko et al., 2012).

Of greater relevance to the present study is the question of how initial language processing shapes false and true memory representations in older adults. With respect to false memory, many previous studies have investigated age-related differences in the so-called DRM paradigm (e.g., coined after Deese, 1959; Roediger & McDermott, 1995). DRM studies present participants with semantically coherent word lists (e.g., "bed, rest, tired, awake, dream, snooze, …"), in which all words are semantically related to a critical lure word (e.g., "sleep"). A common finding in these studies is that older adults often show larger false recognition and recall rates for lure words than is the case for younger adults (e.g., Balota et al., 1999; Norman & Schacter, 1997; Tun et al., 1998, for reviews, see Chang & Brainerd, 2021; Gallo, 2010). Such findings indicate that older adults are less likely to encode and retrieve item-specific information, that is, they overrely on semantic–relational processing, at the expense of encoding and maintaining information that is unique to single items (e.g., Koutstaal & Schacter, 1997). Interestingly, older adults also tend to overestimate the veracity of their memory contents (e.g., Dodson et al., 2007), by issuing a large proportion of false-memory judgments with high confidence (e.g., Norman & Schacter, 1997). This indicates that aging individuals not only misremember semantically related information, but they also have difficulty correctly gauging the contents of their memory (e.g., Greene et al., 2024).

Age differences in false remembering have also been investigated by studies using sentential materials, as opposed to word lists (e.g., Failes et al., 2020; Gunter et al., 1992, 1995; Hartman & Hasher, 1991; Matzen & Benjamin, 2013), even though these studies are far less consistent in their results, compared to previous DRM studies.

For example, in an early study by Gunter et al. (1992), younger and older adults showed comparable rates of false remembering for predictable lures (e.g., “chips”) after encoding constraining sentences that ended with an unpredictable word (e.g., “She ordered fish and … dogs”). However, that and other studies may have inadequately quantified age-related differences in false memory, by only taking into account absolute rates of false remembering. Such an approach is problematic because older adults often show larger response bias in recognition tests than younger adults, that is, a greater likelihood to endorse any kind of material as old (e.g., Fraundorf et al., 2019).

A clear insight that emerges from the extant literature is that memory declines are not universal in aging but malleable by encoding strategies (e.g., Payne et al., 2014; Steen-Baker et al., 2017; Stine & Hindman, 1994; Stine-Morrow et al., 1996, 2006). For example, Stine and Hindman (1994) asked groups of younger and older adults to read sentences that varied in propositional density for immediate recall. Even though older adults recalled fewer propositions than younger adults did, their recall performance for the more difficult dense sentences was comparable to their recall performance for less difficult sentences, after having allocated more time during initial encoding to read these more difficult sentences. Hence, aging readers in that study allocated more time to cognitively organize and encode more complex materials in a possibly strategic manner to benefit their subsequent memory (i.e., a task trade-off). Converging evidence for the notion of self-regulated language processing in aging comes from studies showing that age increases in false remembering are attenuated in paradigms that direct attention to item-specific information, for example, when study words are presented with accompanying pictures or when older adults are instructed about the nature of the false-memory manipulation (e.g., Arndt & Reder, 2003; Koutstaal et al., 1999; Schacter et al., 1999; Thomas & Sommers, 2005; among others). Hence, aging readers self-regulate their sentence and text processing (e.g., Payne et al., 2014; Stine & Hindman, 1994; Stine-Morrow et al., 1996, 2006, 2008).

In contrast to the false-memory literature, the extant literature on recognition for sentence content (i.e., true memory) is more uniform in showing that older adults’ memory may be preferentially driven by contextually matching compared to mismatching information (i.e., a schema congruency effect in older adults; e.g., Badham et al., 2012; Brod et al., 2013; Brod & Shing, 2019; Klever et al., 2023; Skinner & Price, 2019). For example, studies using sentence materials found that memory retrieval of single words (e.g., “spray”) was preferentially improved in older adults when the words were embedded in a contextually predictive context (e.g., “Kill the bugs with spray” vs. “Mary considered the spray”; e.g., Gordon-Salant & Fitzgibbons, 1997; also see McCoy et al., 2005). Hence, older adults’ memory might rely more on schemas. Some researchers believe that this reliance may be a consequence of older adults’ lifelong-accumulated knowledge base, which may result in a diminished capacity to process novel information (e.g., Shing et al., 2023). Others believe that the schema congruency effect may be driven by a preference to encode and review information that is known and familiar, rather than new and possibly challenging (e.g., Carstensen et al., 1999; Mather, 2004).

## The Present Study

We investigated age differences during initial reading of predictable and unpredictable sentences and how these affect the memory representations that younger and older adults create and maintain (e.g., Kintsch et al., 1990). German participants who varied in age read semantically rich, constraining sentences that continued in a more or less predictable word. In a (surprise) word recognition memory task that followed shortly after, participants were presented with single words and had to indicate whether the words were old or new and how confident they were with their response (i.e., using a 4-point response scale for *sure new*, *maybe new*, *maybe old*, and *sure old*). Participants were not only presented with old predictable and unpredictable words, which allowed us to examine age-related differences in reliance on schema versus prediction error (e.g., Shing et al., 2023), but they were also presented with predictable but not actually presented words (i.e., “lures”), which allowed us to investigate age-differences in false memory.

For self-paced reading, we expected to find a reduced predictability effect in older compared to younger adults, consistent with the idea that older adults use sentential context less readily than younger adults (e.g., Noh & Stine-Morrow, 2009). For recognition memory, we predicted that aging should increase reliance on semantic gist (e.g., Koutstaal & Schacter, 1997). This means that we expected to find a larger false-memory effect for older than younger participants when additionally taking into account response bias (i.e., more false alarms to lures than to new nouns). For *true* (as opposed to false) memory, we expected to find that older adults should be more likely to remember schema-congruent information. Finally, in line with the notion of self-regulated sentence processing (e.g., Payne et al., 2014; Stine & Hindman, 1994), we expected to find a pattern of results where some older readers achieve relatively high levels of memory performance, in part by allocating disproportionately more time to text-based processes.

# Method

## Participants

The study was run online (see below for details) in the time period from July until October 2021. A total of 150 German participants was recruited through social media and online platforms. Three subjects were excluded before analysis (i.e., two subjects for reading/writing disabilities, another subject for pausing the study for more than 30 min between encoding and retrieval tasks). Four subjects were excluded due to below 70% accuracy on the self-paced reading (SPR) comprehension questions. Informed consent was obtained from all participants. Ethics approval was granted through the German Research Foundation, research project A5, *The Role of Language Experience and Surprisal In Learning and Memory* (Collaborative Research Center 1102, project-ID 232722074).

The final sample of participants consisted of 99 younger adults ($M_{age}$ = 27 years), recruited through Prolific ($n$ = 49, 32 males, 17 females, $M_{age}$ = 28 years, age range = 18–40) and an online platform which advertises experimental studies to Psychology students in exchange for course credit ($n$ = 50, 16 males, 34 females, $M_{age}$ = 25 years, age range = 18–40). The older group of participants was recruited through a Facebook website that promotes recent findings in Psychology research to a lay audience. A total of 48 older adults (39 males, nine females, $M_{age}$ = 60 years, age range = 49–77) was willing to participate. The final sample of participants reported no history of reading or writing disabilities and/or taking neuropsychiatric medication at the time of testing. All participants were White Europeans. Table 1 shows participants' education background.

## Materials and Tasks

### *SPR Task*

Materials for the SPR task consisted of 44 constraining German sentence frames (e.g., "At the hospital, the nurse stitches up quickly the …"), which continued in a relatively predictable (e.g., "injury") and relatively less predictable (but somewhat plausible) noun (e.g., "pants"). All sentence frames and their nouns were taken from an earlier study using German materials and were normed for predictability in both younger and older adults (see Haeuser & Kray 2022b). Cloze probability rates for predictable nouns did not differ between younger ($M = 0.78$) and older adults ($M = 0.76$), $t(174) = 0.18$, $p = .86$. All sentences were also normed for plausibility in younger adults (see Haeuser & Kray 2022b). Table 2 shows descriptive statistics for predictable and unpredictable nouns used in the experiment.

All predictable and unpredictable sentences contained two- or three-word sentence continuations after the noun that continued the sentence in a plausible fashion (e.g., "im Zimmer nebenan," "in the room next door"; "um die Ecke," "around the corner"; "am Morgen," "in the morning"), in order to allow for the measurement of spillover effects that frequently occur in self-paced sentence reading (e.g., Witzel et al., 2012). Note that the spillover region was identical for predictable and unpredictable versions of an item.

The experimental sentences were presented in two lists using a Latin square design, such that one participant was presented with the predictable or unpredictable version of an item during the experiment but not both. Each list additionally contained 25 moderately constraining filler sentences from the Potsdam sentence corpus, which were included to ensure that participants continued to generate predictions in the course of the experiment (i.e., despite frequently having predictions disconfirmed; see e.g., Delaney-Busch et al., 2019; Fine et al., 2013). All experimental (but not filler) sentences were followed by yes/no comprehension questions, inserted to ensure that participants were reading for comprehension. Comprehension questions were identical for predictable and unpredictable versions of an item and mostly probed information in the sentence context.

### *Noun Recognition Task*

Materials for the recognition task consisted of 88 old nouns (i.e., previously seen predictable and unpredictable nouns in SPR), 44 new nouns (i.e., nouns not previously seen in SPR), and 44 lure nouns (i.e., nouns previously predicted but not seen). A single participant saw a total of 110 nouns in the recognition task, that is 44 previously seen old nouns (22 predictable, 22 unpredictable), 44 new nouns, and 22 lures (i.e., the lures from the 22 unpredictable sentences seen earlier during encoding). New nouns were obtained from a subset of German words in the movie subtitle database (i.e., SUBTLEX; Brysbaert et al., 2011), taking advantage of the fact that the first letter in all German nouns is capitalized. In a previous step, all nouns used in the experiment (i.e., all capitalized words in the self-paced reading task and its practice session) had been excluded from that noun list. The experimenters then chose new nouns that roughly fell in the same frequency range as the old nouns. Old, new, and lure nouns were matched with respect to frequency, $F(2, 172) = 0.83$, $p > .05$, but differed from one another in length, $F(2, 173) = 3.66$, $p < .05$, such that lure and old words were, on average, shorter than new words, Welch's $t(71.52) = -2.78$, $p = .01$, and $t(171.15) = 3.01$, $p = .01$, respectively. Note that our statistical analysis controlled for these differences. Table 3 shows descriptive statistics for nouns used over the three conditions: old, new, and lure.

All recognition nouns were presented on two experimental lists. Only those participants who got to see the unpredictable version of an experimental sentence during SPR were presented with the predictable lure in the recognition task. Participants who got to see the predictable version of a sentence during self-paced reading were presented with the predictable lure, an old word during recognition.

## Procedure

The study was run online using the experimental platform LabVanced (Finger et al., 2017), a JavaScript web application that allows for online implementation of behavioral research, offering similar precision and data quality as traditional lab-based testing (e.g., Finger et al., 2017). Participation was only possible through a PC (Windows or Linux), Mac, or tablet; cell phone or iPhone participation was disabled. The study consisted of three major sections. The first section was a noncumulative word-by-word self-paced reading task (∼15–20 min), followed by a brief retention interval (∼5 min), in which participants solved simple math problems. The math task served the primary purpose of clearing participant's short-term memory before they proceeded to the noun recognition memory task. The third section of the experiment was the noun recognition task (∼15 min). In its initial stages the study also included a verbal fluency test (semantic and phonemic), which we inserted to test vocabulary levels and linguistic control. The verbal fluency task was removed after the study aired because most of the older participants refused to consent to their voice being recorded when entering the study.

**Table 1**
*Participants' Education Background*

| Group | No higher degree | High school degree | University degree |
|---|---|---|---|
| Psychology students | 0 | 49 | 1 |
| Prolific sample | 5 | 13 | 31 |
| Older adults | 17 | 14 | 17 |

*Note.* We are using "high school degree" as an approximation to refer to participants who had completed the German *Abitur*, that is, the entry-level exam that qualifies people to study at German universities.

**Table 2**

*Descriptive Means (and Standard Deviation) of Critical Nouns in the Self-Paced Reading Task*

| Condition | Length (*N* characters) | Frequency (SUBTLEX-DE) | Cloze probability younger adults | Cloze probability older adults | Plausibility (scale = 1–7) | Concreteness (scale = 1–9) |
|---|---|---|---|---|---|---|
| Predictable | 4.59 (2.60) | 4.11 (1.29) | 0.78 (.16) | 0.76 (.18) | 6.57 (0.33) | 2.13 (0.95) |
| Unpredictable | 4.68 (2.58) | 4.06 (1.35) | <.001 | <.001 | 3.88 (1.32) | 2.49 (1.29) |

*Note.* Frequency estimates are log-per million values (Brysbaert et al., 2011). Concreteness ratings (with a rating of 1 meaning *very abstract* and a rating of 9 meaning *very concrete*) were obtained from the Leipzig Affective Young Adult Norms for German (Kanske & Kotz, 2010) and were available for 62 out of the 88 nouns used here. Plausibility ratings (1 = *plausible*, 7 = *implausible*) were known through an earlier young-adult study in our lab (Haeuser & Kray 2022b).

In the self-paced reading task, participants read the experimental sentences on a screen word by word. Each word was only displayed once and was not replaced by dashes later on (i.e., there was no mask, no moving-window reading). Each word was presented in the center of a white screen using Lucinda 18-point font and stayed on the screen until participants pushed the space bar, which revealed the next word in the sentence. Participants were instructed to read the sentences as quickly and thoroughly as possible and to answer all comprehension questions as accurately as possible by pushing the "S" (yes, correct) and "L" (no, incorrect) keys on the keyboard. The experimental task was preceded by 10 practice sentences, which made sure that participants could get used to the word-by-word reading task.

In the noun recognition memory task, participants saw single nouns appear on the screen in a central position using a Lucinda 18-point font, separated by a central fixation cross displayed for 5,000 ms. Participants were instructed to hover their middle and index fingers of each hand over the S, D and J, K keys on the keyboard. They were instructed to press the S or D key whenever they thought that a word was "sure new" or "maybe new" and to press the J or K key for all "maybe old" and "sure old" words (i.e., S = sure new, D = maybe new, J = maybe old, K = sure old). Figure 1 shows the response scale.

Participants were instructed that they did not need to memorize the keys combinations, because every trial contained a schematic display of the response options (SD, JK) and what they represented, at the bottom of the screen. Participants were instructed to respond to each word as quickly and accurately as possible. About two thirds of all participants completed the experiment within 24–40 min (average finish time = 36 min).

## Data Analysis

Results for the word-by-word self-paced reading task and subsequent word recognition task are reported in two separate sections; we also report results from models that directly related encoding times to subsequent memory retrieval.

To statistically analyze the reading and recognition data, we ran linear mixed-effects models (Baayen et al., 2008) as specified in the lme4 library (e.g., Bates et al., 2015) in R (R Core Team, 2022; Version 3.6.2), which were run on unaggregated trial-by-trial data. In order to better account for the continuous distribution of ages in our participants (see above), we used age as a continuous variable in all statistical analyses, for which linear mixed-effects models are well-suited (e.g., Baayen et al., 2008; Winter, 2019). Only for graphs, we used a categorical variable for age group, which was based on a median split of the continuous age variable.[1]

All models included random intercepts for subjects and items and were initially fit with the fullest by-item and by-subject random slope structure warranted by the design (i.e., "maximal" model; Barr et al., 2013). In the case of convergence warnings, models were simplified progressively using the *least variance* approach (e.g., Barr et al., 2013); *p* values were calculated using the R package *lmerTest* (Kuznetsova et al., 2017). Predictors that did not contribute substantial variance were removed from the model by means of model comparisons.

[1] We replicated all analyses reported below when using the binary variable for age group.

**Table 3**
*Descriptive Means (and Standard Deviation) of Nouns Used in the Recognition Task*

| Condition | Length (*N* characters) | Frequency (SUBTLEX-DE) | Concreteness (scale = 1–9) |
|---|---|---|---|
| Old | 6.08 (2.16) | 2.63 (0.68) | 2.19 (1.03) |
| New | 7.00 (1.90) | 2.74 (0.02) | 3.09 (2.04) |
| Lure | 5.86 (2.36) | 2.74 (0.72) | 2.13 (0.95) |

## Transparency and Openness

All de-identified data files, analysis scripts (R code), and sentential and recognition materials can be found on this article's project page. We report measures of sentence reading and recognition memory. We exclude data points based on unnaturally short and long reaction times. We exclude participants based on low accuracy rates during sentence reading and abnormally low recognition memory performance. All data exclusions are disseminated in the article.

The study was not preregistered. There were no a priori power simulations, due to the known issues in accurately predicting random effects and slope parameters in linear-mixed effect models (e.g., Brysbaert & Stevens, 2018). According to power standards for aggregated data analysis techniques (e.g., analysis of variance), the present study had 80% power to detect a large effect in recognition memory (i.e., Cohen's $f = 0.51$).

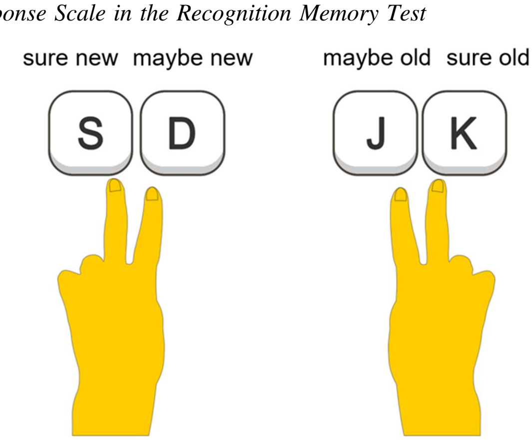
# Results

## Self-Paced Reading

### *Comprehension Accuracy*

We first inspected accuracy rates on the SPR comprehension questions. Four participants had accuracy rates below 70% and, thus, were excluded from further analysis (two subjects from the young-adult group and two from the old-adult group). On average, the remaining participants responded to the comprehension questions very accurately ($M = 0.91$, range = 0.73–1). We subjected the comprehension accuracy data to a 2 (i.e., Older–Younger Adults) × 2 (i.e., Predictable–Unpredictable Items) analysis of variance. The results showed a significant main effect of group, $F(1, 45) = 4.26$, $p = .04$, such that older adults ($M_{accuracy} = 0.90$) responded less accurately to the comprehension questions than younger adults ($M = 0.92$), *Welch's* $t(89.41) = 2.04$, $p = .04$. There was no main effect of predictability, $F(1, 145) = 1.95$, $p = .16$, and no interaction between predictability and age group, $F(1, 145) = 1.13$, $p = .29$. Thus, participants of all ages understood predictable and unpredictable sentences accurately, but older adults seemed to understand and maintain less information from the sentences, compared to the younger adults. We return to this point below.

**Figure 1**
*Response Scale in the Recognition Memory Test*




*Note.* See the online article for the color version of this figure.

### *Reading Times*

The main analysis of the self-paced reading data is available as an R-script on this article's Open Science Framework (OSF) page (file name: SPR_MainAnalysis.R). The critical region for statistical analysis of the reading time (RT) data consisted of the predictable and unpredictable noun, as well as the two words immediately following the noun to catch spillover effects common to word-by-word self-paced reading (i.e., the noun; the first word after the noun, $N + 1$; and the second word after the noun, $N + 2$). We fit separate models for each word in the critical region. The dependent variable was reading times, log-transformed to avoid skewness.[2] Fixed effects were predictability (two levels: predictable vs. unpredictable), sum-coded to allow for interpretation as main effect (cf. Brehm & Alday, 2022; Schad et al., 2020; Winter, 2019), and age, a scaled continuous variable (i.e., centered and standardized), as well as the interaction between predictability and age. We also implemented a control variable for education, given the sample differences in this variable (see Table 1). Education was included as a binary variable (i.e., individuals with lower and higher education, sum-coded with −1 and 1), since preliminary analyses had shown no RT differences between individuals with no higher degree and high school degree, but substantially facilitated reading in individuals with university degree. Additional control predictors in each model consisted of trial number to offset effects of customization (i.e., speed-up effects over time that commonly occur in self-paced reading experiments; e.g., Witzel et al., 2012), word position in the sentence, word frequency (i.e., log-per-million frequency values from the movie subtitle database; Brysbaert et al., 2011), and word length. Initially, we also included a control predictor for scaled reading times of the previous word, but since the inclusion of this variable led to multicollinearity with the predictor *predictability* in models Noun + 1 and Noun + 2, it was dropped. We

[2] We also ran follow-up models on de-logged reading times. The rationale for these models was that age differences often emerge in the rightward tail of an RT distribution, which log-RTs inevitably "hide" (e.g., Jongman et al., 2023; Payne & Federmeier, 2017). Qualitatively though, the de-logged models showed the same effects as the log-models, though the age-related speed up was less pronounced in the de-logged models. We chose to report log-RT models in our main results as they are better suited for our chosen analysis method, that is, linear mixed-effects models.

did not include concreteness as a control variable because concreteness norms were only available for a small subset of our items.

All continuous control predictors, except for frequency (which is log-transformed), were scaled. Before analysis, RT data were trimmed minimally based on visual inspection of the data, by excluding RTs shorter than 200 ms (across words) and longer than 2,500 ms (for the noun; 1,500 ms and 2,000 ms, respectively, for the first and second word after the noun). This procedure affected fewer than 1% of all data points. Table 4 shows final model outputs. Figure 2 shows a partial effects plot.

**Noun.** There was a significant effect for age ($p = .05$; see Table 4), which suggested that, unexpectedly, older individuals read more quickly than younger individuals. The main effect of predictability and the interaction between predictability and age were not significant ($p = .31$ and $p = .60$, respectively).

**Noun + 1.** The model showed a main effect of predictability ($p = .002$), suggesting slowed reading comprehension for unpredictable items, irrespective of age. The main effect of age was not statistically significant ($p = .09$). The interaction between predictability and age was not significant ($p = .19$).

**Noun + 2.** There was a significant effect of predictability ($p < .001$), indicating prolonged reading times for unpredictable over predictable items. The main effect of age was not significant in this model ($p = .23$). The interaction between predictability and age was also not significant ($p = .35$).

In sum, analysis of the self-paced reading data showed two key findings: First, unpredictable words slowed reading comprehension in an age-invariant manner, suggesting that participants of all ages experienced integration difficulties when reading nouns that were unpredictable based on context. As is common in self-paced reading (e.g., Witzel et al., 2012), predictability effects were delayed and did not emerge directly on the critical noun but only on the spillover region. Second, and unexpectedly, older adults seemed to read more quickly than younger adults.

### *Age Differences in Reading Rates Based on Reading Differences in the Young-Adult Sample*

Given the visible differences in reading rates between the two young-adult samples recruited for this study (see Figure 2), we investigated whether our findings were modulated by systematic differences in reading rates between participants recruited over Prolific and Psychology students (the R code for this analysis is included in the OSF page of this article; file name: SPR_FollowUp.R).[3] We reanalyzed the self-paced reading data with a grouping variable for the three participant groups, that is, the student sample, young-adult Prolific sample and older sample. Models identical to those in the main Results section were run separately for each word in the critical region, but instead of implementing the scaled continuous variable for age, we now used a three-level factor for participant group (Helmert coded) in which we compared, in a first contrast, the Prolific sample against the Psychology student sample, and, in a second contrast, the older sample against the Prolific sample.

The results are easily summarized: First, in all models, younger adults recruited over Prolific read considerably more quickly than the sample of the young-adult Psychology students (all $p$'s $< .05$; see Figure 2). Second, when we selectively compared the older sample to the Prolific sample, the older adult speed up in reading times disappeared: Across models, there were no differences in reading rates between the older adults and the sample of Prolific workers (all $p$'s $> .20$; see Figure 2).

In sum, it seems that the unexpected speed-up in older adults' reading times was driven by a relatively more slowly reading subsample of younger adults. A remaining possibility is that older adults gained their initial reading speed at the expense of accurately encoding sentences into memory. A first indication attesting to this possibility are the SPR sentence comprehension questions (see above), which had shown less accurate comprehension in older compared to younger adults. To garner more empirical support for this hypothesis, we now to turn to the results from the recognition memory test.

## Noun Recognition Memory

The recognition memory results for the two groups of younger adults (i.e., Prolific workers and Psychology students) did not differ from one another statistically (except where specified; see below), so we combined the data from all age samples, again using the scaled continuous variable for age in models on recognition memory. For full disclosure, plots and summary statistics are reported for the three samples. The analysis files are available on this article's OSF repository (file name: Recognition_MainAnalysis.R).

Four participants (aged 30, 52, 63, and 76) were excluded from all following analyses because they had a higher false alarm than hit rate. Our main analysis below examines false and true recognition memory by means of effect coding, which takes into account differences in response bias. Age-related differences with respect to discrimination between old and new items is included in this analysis. Results from an additional analysis that examined $d'$ and the criterion shift measure $c$ aligned with our main results and are presented in the Supplemental Material of the article.

To statistically analyze the recognition data, we ran generalized linear mixed-effects models. Prior to analysis, recognition judgments faster than 200 ms and slower than 10,000 ms were removed, which affected less than 1% of all data points. The measured variable were trial-by-trial old judgments, that is, whether a word was judged as old or new (coded in 1s and 0s). For previously seen (old) nouns, old judgments reflect hits. For lures and new items, old judgments reflect false alarms. Fixed effects in the model were *word type* (four levels: old-predictable, old-unpredictable, new, and lure; see below for contrast coding), the scaled (i.e., centered and standardized) continuous variable for age, and confidence (low vs. high, sum coded with −1 and +1). The model also contained scaled control predictors for word length, log-per-million frequency from the movie subtitle database (Brysbaert et al., 2011), and trial number. We did not include education as a control variable in this analysis, because it did not account for substantial variance in the data.[4] To facilitate model convergence, we used the Bobyqua optimizer. For the predictor *word type*, we set up three contrasts, following the procedure described in Schad et al. (2020).

[3] We also investigated whether age effects in reading rates were additionally modulated by educational attainment. This analysis showed no effects and is presented in the online supplement of this article.

[4] We investigated if recognition rates were additionally modulated by individual differences in education and report these analyses in the online supplement of the article. Overall, the results showed no effects of interest.

**Table 4**
*Effects of Predictability and Age on Reading Times in the Critical Region*

| | Noun | | | Noun + 1 | | | Noun + 2 | | |
|---|---|---|---|---|---|---|---|---|---|
| Predictor | *b* | *SE* | *t* | *b* | *SE* | *t* | *b* | *SE* | *t* |
| Fixed effects | | | | | | | | | |
| Predictability | 0.004 | 0.004 | 1.02 | 0.01 | 0.004 | 3.03 | 0.02 | 0.004 | 4.64 |
| (Continuous) age | −0.06 | 0.03 | −2.01 | −0.04 | 0.02 | −1.70 | −0.03 | 0.03 | −1.30 |
| Frequency | −0.01 | 0.01 | −0.62 | −0.003 | 0.02 | −0.21 | | | |
| Length | 0.02 | 0.01 | 4.38 | −0.001 | 0.02 | −0.21 | 0.08 | 0.07 | 10.56 |
| Word position | −0.001 | 0.002 | −0.01 | 0.01 | 0.003 | 3.48 | 0.01 | 0.003 | 2.21 |
| Trial number | −0.06 | 0.004 | −13.75 | −0.05 | 0.004 | −11.77 | −0.05 | 0.004 | −12.01 |
| Education | −0.05 | 0.03 | −1.98 | −0.06 | 0.02 | −2.54 | −0.07 | 0.03 | −2.78 |
| Predictability × Age | 0.002 | 0.004 | 0.53 | −0.01 | 0.004 | −1.32 | −0.003 | 0.003 | −0.93 |
| Random effects (variance) | | | | | | | | | |
| Subject | | 0.11 | | | 0.07 | | | 0.08 | |
| Predictability | | 0.001 | | | 0.001 | | | 0.001 | |
| Item | | 0.001 | | | 0.002 | | | 0.002 | |
| Predictability | | [a] | | | [a] | | | [a] | |
| Age | | 0.001 | | | 0.001 | | | 0.001 | |
| Predictability × Age | | [a] | | | [a] | | | [a] | |

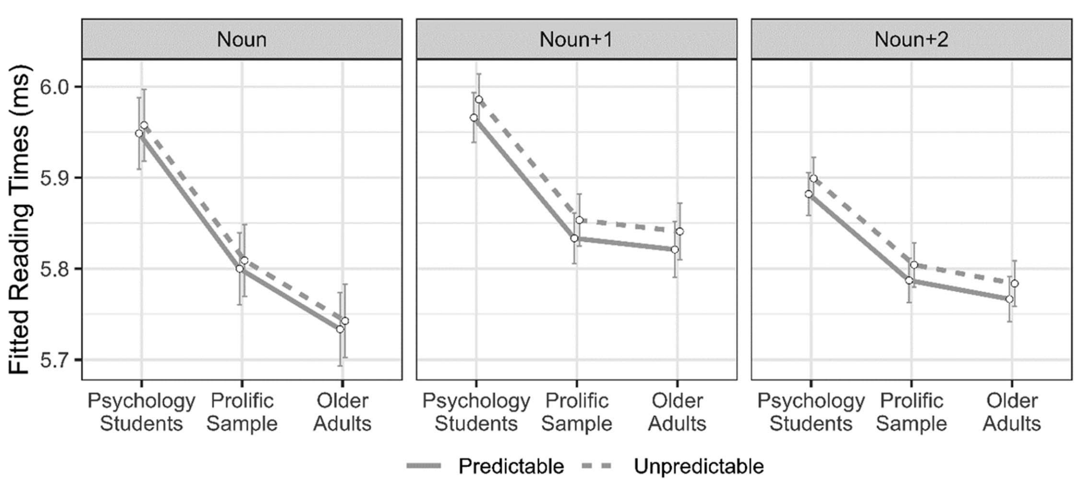
*Note.* The predictor frequency was removed from the Noun + 2 model, as frequency estimates for the words in this region were only available for half of our experimental items. *t* values larger than 2.0 are significant at the .05 level. *SE* = standard error.
[a] Indicate predictors that were removed from the model due to convergence issues.

The first contrast *old-versus-new* (not part of our initially motivated hypotheses) served as a manipulation check, as it measured whether our participants successfully discriminated between old and new items. If discrimination is successful, hit rates to old items should be higher than false-alarm rates to new items. To the extent that younger and older adults show different discrimination rates for low- and high-confidence judgments (e.g., Dodson et al., 2007), this effect should be modulated by age and confidence.

The second contrast *lure-versus-new* measured false-memory effects by comparing false alarms for lures against false alarms to new items. If predictable words elicit false-recognition memory, false alarm rates to lures should be consistently higher than those to new words. In line with previous studies using the DRM paradigm (e.g., Norman & Schacter, 1997), this effect may also be modulated by age and confidence, such that older adults demonstrate larger false-memory effects in high-confidence judgments.

The third contrast (i.e., *unpredictable vs. predictable*) measured effects of schema congruency versus prediction error, by comparing hit rates for predictable nouns to those of unpredictable nouns. If learning is preferentially driven by schema congruency in older adults, aging might increase hit rates for predictable nouns.

**Figure 2**
*Partial-Effects Plot of Reading Times (ms) in the Self-Paced Reading Task*




*Note.* The graph shows age as a categorical predictor with three levels corresponding to the three participant groups; all statistical models were run with the scaled continuous variable for age. The graph also shows untransformed (de-logged) reading times. All statistical analyses were run on log-transformed reading times.

### *Results*

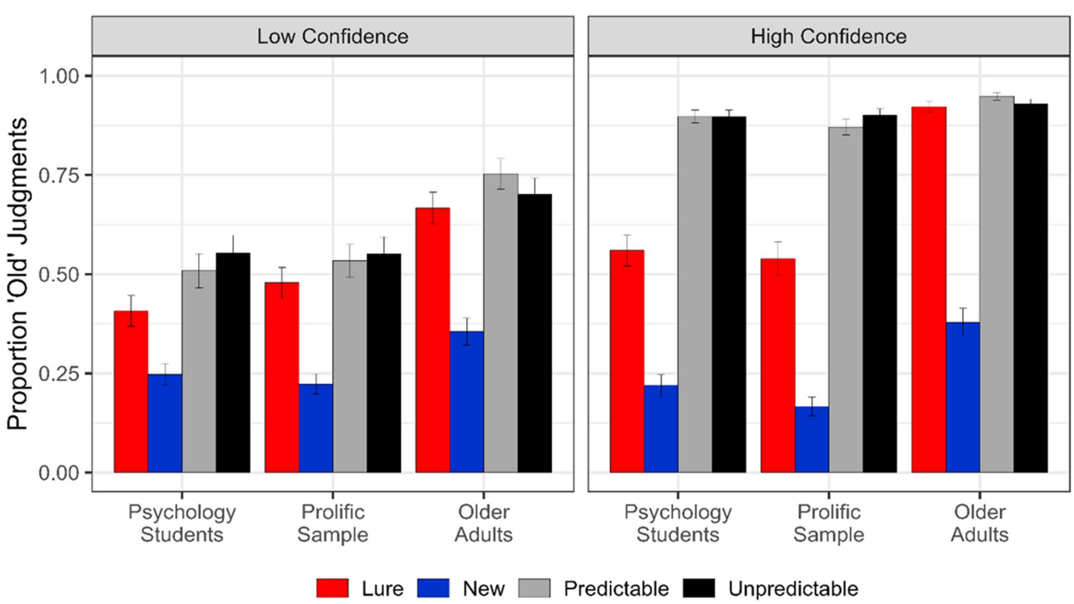
The final model converged with by-subject and by-noun random intercepts and by-subject random slopes for confidence. A partial effects plot of the model is presented in Figure 3.

For the contrast old-versus-new (i.e., true recognition memory): The model showed a significant three-way interaction between the contrast old-versus-new, age, and confidence ($b = -0.11$, $SE = 0.05$, $z = -2.16$, $p = .03$). Figure 3 indicates that this effect was predominantly driven by high-confidence judgments, where older adults showed less successful discrimination between old and new items. Follow-up models that split the data by low- and high-recognition judgments confirmed this impression, though the interaction between old-versus-new and scaled age did not quite reach statistical significance in the model for the high-confidence judgments ($b = -0.12$, $SE = 0.08$, $z = -1.51$, $p = .13$; interaction *old-vs.-new* and age in low-confidence judgments: $b = 0.09$, $SE = 0.06$, $z = 1.49$, $p = .14$; high-confidence judgments). There were no significant differences in old–new discrimination between the two young-adult samples ($z = 1.71$, $p = .09$). Altogether, these findings indicate that participants of all age groups discriminated relatively successfully between old and new words, even though aging reduced successful discrimination in high-confidence judgments (for converging results using $d'$ and $c$, see Online Supplement).

For the contrast *lure-versus-new* (i.e., false-recognition memory), the model showed a significant three-way interaction between *lure-versus-new*, age, and confidence ($b = 0.30$, $SE = 0.06$, $z = 5.24$, $p < .001$). Follow-up models, in which we split data by confidence, suggested that the effect was primarily driven by high-confidence memory judgments: Specifically, aging increased the false-memory effect in high-confidence judgments (interaction between *lure-vs.-new* and age in high-confidence judgments: $b = 0.77$, $SE = 0.09$, $z = 8.11$, $p < .001$), whereas in low-confidence judgments, the main effect of *lure-versus-new* ($b = 0.90$, $SE = 0.11$, $z = 8.45$, $p < .001$) was not additionally modulated by age (interaction between *lure-vs.-new* and age in low-confidence judgments: $b = 0.05$, $SE = 0.07$, $z = 0.76$, $p = .45$). The young-adult Prolific sample showed a larger false-memory effect than the Psychology student sample (interaction *lure-vs.-new* and *Prolific vs. Psych students:* $b = 0.35$, $SE = 0.13$, $z = 2.83$, $p < .01$; note that absolute false alarm rates to lures were relatively similar in the two young-adult samples, see Figure 3). Hence, participants of all ages experienced a 0strong false-memory effect for predictable lure. However, aging increased the false-memory effect selectively in high-confidence judgments, which could indicate that older adults are more prone to high-confidence memory intrusions, in line with previous research (e.g., Dodson et al., 2007). We confirmed these impressions in a follow-up analysis that compared age-related differences in the proportions of responses that participants allocated to each memory bin (i.e., sure new, maybe new, maybe old, sure old; see Supplemental Material of this article). In that analysis, older adults allocated proportionally *more* old responses than younger adults to lures with high confidence (i.e., sure old judgments), but they allocated *fewer* low-confidence maybe old responses. This demonstrates an age-related shift from low- to high-confidence false-memory judgments.

For the contrast *old-unpredictable versus old-predictable*, the model showed a significant two-way interaction between old-unpredictable versus old-predictable and age ($b = -0.21$, $SE = 0.08$, $z = -2.66$, $p < .01$). This effect was not additionally modulated by confidence (three-way interaction between old-unpredictable vs. old-predictable, age, and confidence: $b = 0.002$, $SE = 0.08$, $z = 0.03$, $p = .98$). This shows that, irrespective of confidence, aging increased hit rates for previously seen predictable, compared to unpredictable,

**Figure 3**
*Partial Effects Plot of Old Judgments in the Recognition Memory Task*




*Note.* All statistical analyses were run using the scaled continuous variable for age. Error bars represent standard error (*SE*) Lure = previously predictable but not presented noun; New = new noun; Predictable/Unpredictable = previously presented predictable or unpredictable noun. See the online article for the color version of this figure.

nouns—in other words, there was a schema congruency effect in older but not younger adults. Of note, we ran two follow-up models on old items, in order to investigate whether the schema congruency effect in older adults was modulated by previous reading times and, possibly, comprehension accuracy rates during SPR. The results showed that neither was the case (main effects: $z = 1.73$, $p = .83$, and $z = 1.19$, $p = .24$, respectively; three-way interaction between reading times/comprehension accuracy and condition and age: $z = -0.59$, $p = .55$ and $z = 0.53$, $p = .61$, respectively).

In sum, the data from the recognition memory test showed three key findings. First, participants of all ages showed a false-memory effect for predictable lures, across confidence judgments (see Figure 3), but aging increased rates of false remembering specifically in high-confidence judgments, where older adults' rates of false remembering approached their rates of true remembering. In line with this, aging also attenuated successful discrimination between old and new items in high-confidence judgments. Third, irrespective of confidence, aging increased hit rates for predictable versus unpredictable words.

## Exploratory Analysis: Relating Previous Reading Times to Recognition Memory

Our main analysis on recognition memory did not allow us to include previous encoding RTs during self-paced reading (as a predictor, because encoding RTs are not specified for new words). To more directly relate encoding RTs to memory performance, we present an exploratory analysis that investigated whether false-recognition memory rates were related to previous reading times of prediction-disconfirming sentence continuations. Specifically, we investigated whether false-recognition judgments to lures were modified by the speed with which participants read the prediction-disconfirming critical region for that item during encoding. To use an example, we speculated that prolonged reading of "At the hospital the nurse stitches up quickly the *girl in the* room next door" may more readily suppress false memories for "injury." We expected that older adults may be particularly likely to benefit from longer reading times as they are known to more readily self-regulate their encoding (Stine & Hindman, 1994). We operationalized encoding RTs as a composite score, which was specified as the average RT of the spillover region after the noun (i.e., Noun + 1 and Noun + 2 words), as only the spillover region had shown clear predictability effects during reading.[5] In the event of missing reading time values (i.e., reading time observations that were removed as outliers), the composite score was computed as the mean of the remaining cells (see R-File Exploratory_Analysis.R on OSF). The outcome variable in the model were old judgments to predictable lures. Predictor variables were scaled age, and the scaled composite score for RTs of the prediction-disconfirming sentence continuation for the same item, including the interaction between the two.[6]

Critically, the interaction between age and previous reading speed was significant in the model ($b = -0.14$, $SE = 0.05$, $z = -2.70$, $p < .01$). Figure 4 suggests that, whereas false-memory rates in younger adults were not apparently modulated by previous reading times (see panels for Psychology students and Prolific workers), false-memory rates in older adults were: Those older adults who read the unpredictable sentence continuations more slowly during self-paced reading showed reduced rates of false remembering for predictable lures. Follow-up models, in which we split the data by age group, confirmed these visual impressions: Previous reading times were a significant predictor of false-memory rates in older, but not younger, adults ($b = -0.24$, $SE = 0.13$, $z = -1.93$, $p = .05$, and $b = -0.03$, $SE = 0.06$, $z = -0.53$, $p = .59$, respectively). Taken together, this suggests that false remembering of lures was attenuated in older adults who read prediction-disconfirming sentence continuations more slowly.

## Summary of Findings From Self-Paced Reading and Word Recognition Memory

During self-paced reading, participants of all ages incurred a processing cost when reading unpredictable sentence continuations. Crucially, older adults read the SPR sentences as quickly as the young-adult Prolific sample, and even more quickly when compared to the slow-reading Psychology student sample. However, the older participants' initial reading speed seemed to come at the expense of encoding sentences correctly into memory: Older adults maintained less information about the sentences in short-term memory (as illustrated by their reduced accuracy on the SPR comprehension questions) and also showed declines in long-term recognition memory. Not only did older participants show a false-memory effect for predictable lures that was nearly twice the size of the false-memory effect found for younger participants, but older adults also discriminated less successfully between previously seen, old, and previously not seen, new, words. Both findings were more likely to emerge in high-confidence judgments, and in line with this, older adults issued proportionally more high-confidence old judgments than younger adults did. This indicates that, in addition to understanding the sentences less distinctly than younger adults, older participants were more prone to misjudge the veracity of their memory contents. Altogether, these findings could indicate that older adults gained their initial reading speed at the expense of accurately encoding sentences into memory, which resulted in memory declines in both short- and long-term. However, those older participants, who spent more time reading unpredictable sentence continuations showed lower levels of false memory for predictable lures. Finally, and in line with previous findings, older adults showed a schema congruency effect, that is, better recognition memory for predictable compared to unpredictable words.

# Discussion

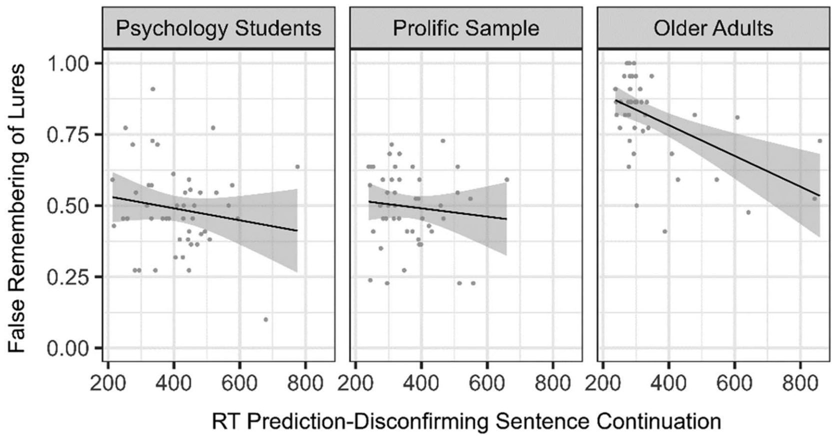
We investigated age differences during initial reading of predictable and unpredictable sentences and how these affect the memory representations that younger and older adults create and maintain (e.g., Kintsch et al., 1990). German participants of varying ages read predictable and unpredictable sentences for comprehension. A subsequent recognition memory task tested participants' recognition memory for words that may have been predicted during reading but were not actually presented (i.e., false-recognition memory for lures). Their recognition memory was also probed for previously seen (i.e., old) predictable and unpredictable words (i.e.,

[5] We also ran a follow-up model with a composite score that included all words in the critical region (i.e., noun, Noun + 1 and Noun + 2). That model showed the same results as reported below.

[6] We dropped confidence from this analysis because the three-way interaction between age, confidence, and encoding RT did not reach statistical significance ($z = 0.51$, $p = .61$); neither did the two-way interaction between encoding RT and confidence ($z = -0.75$, $p = .45$).

**Figure 4**

*Reading Times of Prediction-Disconfirming Sentence Continuations Plotted Against False-Memory Rates for Lures*




*Note*. The figure was generated by aggregating data over items; one point is one subject. The statistical analysis was run on the by-item data. RT = reading times.

true recognition memory) and how confident participants were with their recognition judgments. Overall, our results illuminate a complex pattern of age-related decline and reserve. Older adults encoded fewer propositional details of sentences during initial reading, and they showed poorer memory discrimination and strong increases in gist-wise false remembering later on. At the same time, allocating more time during initial reading of sentences buffered some of the memory deficits in old age. We unpack our findings for younger and older adults, as well as their implications, in greater detail below.

## How Sentence Comprehension Affects Recognition Memory in Younger Adults

The data in the present study clearly show that sentence comprehension has effects that extend beyond the immediate moment. Specifically, we found that sentence comprehension can elicit false remembering: Participants were more likely to false alarm to words that were congruent with, and predictable from (e.g., "lawn"), previous sentence context (e.g., "Yesterday, our friendly neighbor mowed for us the courtyard"). This kind of "false memory" effect was much larger than would be expected based on baseline levels of false remembering alone, as reflected in incorrect endorsements of new words. This aspect of our findings confirms and extends previous research on English-speaking young-adult individuals (e.g., Chung & Federmeier, 2023; Hubbard et al., 2019; Rich & Harris, 2021; Rommers & Federmeier, 2018a) and allows for at least two interpretations regarding the cognitive mechanisms that give rise to this effect. One possibility is that false-recognition memory is caused by association or prediction during sentence comprehension, that is, participants preactivated predictable words when reading the sentence context. The predicted words then remained active in memory and elicited false remembering during the subsequent recognition task (e.g., Hubbard et al., 2019; Rommers & Federmeier, 2018a). Another possibility is that the false-memory effect is generated more downstream during word recognition, where the predictable lure reinstates the sentence context by means of semantic association, eliciting familiarity and reminding participants of the sentence they had read previously. Our current data cannot adjudicate between these two accounts. However, we do note that one aspect of our young-adult data seems inconsistent with a strong "prediction" interpretation of the false-memory effect: The fact that there was no apparent relationship between the speed with which younger adults read unpredictable items and their false-memory rates later on. To the extent that prolonged reading times reflect integration of unpredictable nouns into the sentence context and possibly discarding of initial predictions, one would have expected such a relationship to surface. We are currently running follow-up experiments that will hopefully help adjudicate between these two interpretations of the false-memory effect.

A second goal of the present study was to investigate what promotes successful recognition memory—mismatch or congruency with accumulated knowledge about the world (i.e., schema congruency vs. prediction error; Clark, 2013; Fitz & Chang, 2019; Henson & Gagnepain, 2010). However, the young-adult data in the present study showed no clear effect in either direction. A possible explanation could be that behavioral recognition tasks (such as the one used here) may not be sufficiently sensitive to test the potentially subtle effects elicited by schema congruency versus prediction error in young-adult subjects. For example, a recent study on English-speaking young adult participants found clear predictability mismatch effects in a time-sensitive measure such as event-related potentials but not in a behavioral recognition memory task (Rommers & Federmeier, 2018b). Another possible explanation could be that younger adults require very large "pop-up" or novelty effects in order to facilitate learning, such as severe violations of world-based and event-based plausibility, as has been reported in the literature (e.g., DeLong et al., 2014; Haeuser & Kray, 2022b; Kuperberg et al., 2020; Van de Meerendonk et al., 2010). The prediction-incongruent nouns in the present may have been too subtle (and too plausible) to elicit such kind of an effect.

### Age-Related Differences in Sentence Comprehension and Recognition Memory

Much like in earlier research, older adults in the present study slowed down when reading unpredictable sentence continuations. The processing costs associated with encountering unpredictable sentence continuations were relatively similar for younger and older adults, which supports earlier behavioral research showing that older adults are able to use constraining sentence contexts to facilitate word processing (e.g., Light et al., 1991, 1992; Stine-Morrow et al., 1996). A surprising finding in this study was that older adults read as quickly, if not more quickly, than younger adults. However, aspects of our data show that the unexpected speed-up in older participants may have come at the cost of accurately encoding sentences into memory. Not only did older adults show less accurate comprehension of the sentences, as indicated by the behavioral comprehension questions that followed at sentence offset, but older adults also showed poorer long-term memory performance than younger adults, both with respect to an increase in false memory for predictable lures and with respect to successful discrimination between old and new information. Hence, older participants achieved younglike performance during initial reading of sentences, but their speed seemed to come at the expense of accurately encoding sentences into memory. In line with this, those older adults who read prediction-disconfirming sentence continuations more slowly showed attenuated rates of false remembering—either because they allocated more time to build message-level information from unpredictable sentences (i.e., akin to showing a stronger prediction mismatch effect) or because they generally attended to the sentences more carefully (Craik & Lockhart, 1972). Our current data do not fully allow us to adjudicate between these two accounts: On the one hand, a prediction cost explanation seems unlikely, since a follow-up analysis that estimated false alarm rates to lures depending on per-subject prediction costs (i.e., a subjects' average RTs in the unpredictable minus the predictable condition) showed no effects. On the other hand, a strong "attention" explanation also seems unlikely, since hit rates in the recognition memory task were not apparently modulated by previous reading times.

Our findings from initial reading and recognition memory could indicate a task trade-off between reading and memory performance in older adults, as is reported quite frequently in the aging literature (e.g., Häuser et al., 2018; Stine & Hindman, 1994). We acknowledge that a conscious or deliberate task adaptation strategy seems unlikely, since participants in the present study were not explicitly instructed that their memory would be tested.[7] But the present study did instruct participants to read the initially presented sentences as quickly as possible. It is possible that the older adults took this instruction a little too far, compared to the younger adults—in other words, the older adults read too quickly to maintain detailed information about the sentences. This, in turn, decreased their comprehension accuracy rates during reading, lowered their old–new discrimination rates during recognition, and ultimately, increased rates of false remembering (e.g., Failes et al., 2020; Norman & Schacter, 1997; Tun et al., 1998). Noteworthy here is that previous studies using sentence materials have shown relatively inconsistent effects of associative false remembering in old age (e.g., Gunter et al., 1992; Hartman & Hasher, 1991). However, few of these studies used rich, semantically constraining sentences like we did here. Clearly, the present data show that longer and strongly predictive sentential contexts seem to increase rather than alleviate false-memory rates in aging.

As to the potential reasons for increased rates of false remembering in aging, previous studies have proposed several explanations. One explanation is that older adults rely more on gist-wise encoding in daily situations, such that they maintain the semantic essence of episodes but little item-specific information (e.g., Balota et al., 1999; Koutstaal & Schacter, 1997). This age-related increase in reliance on gist then increases rates of false remembering for meaning-related material. Another possibility is that storing and retrieval operations are, in principle, intact in older adults but that aging individuals "lose" more information along the way due to age-related declines in working memory and episodic monitoring (i.e., cognitive control; Braver & Barch, 2002) or possibly because they fail to inhibit previously generated predictions during reading (e.g., Failes et al., 2020; Rogers et al., 2012).

Interestingly, older adults in the present study were more prone than younger adults to misjudge their memory contents: They allocated more high-confidence old judgments to predictable lures than younger adults, mirroring past studies (e.g., Failes et al., 2020; Tun et al., 1998). In addition, older adults in the present study were also more likely to incorrectly discriminate old and new words in high-confidence judgments and less so in low-confidence judgments. These findings illustrate the paradoxical relationship between memory accuracy and memory confidence that is often found for aging individuals (e.g., Dodson et al., 2007; Greene et al., 2024). Normally, one would expect that individuals who know about their memory declines (such as older adults) are aware that their memories are more impoverished, relative to those of young adults, and that they should take this impoverishment into account by very rarely expressing high confidence. That is, older adults should "downregulate" (Greene et al., 2024) their retrospective confidence ratings, and only express high confidence in situations where they are likely to be accurate. Vice versa, one would expect that high-confidence memory judgments in such individuals, when issued, would likely be correct most of the time. The fact that neither is the case shows that older adults have difficulty calibrating their confidence to their observed memory abilities (e.g., Dodson et al., 2007; Greene et al., 2024). Hence, older adults are poor subjective judges of their own memories.

With respect to recognition memory for predictable versus unpredictable words, the older adults in this study showed better remembering for contextually predictable compared to unpredictable words. This fits with the previous aging literature that has attested to larger schema congruency effects in aging (e.g., Badham et al., 2012; Brod et al., 2013; Klever et al., 2023; Naveh-Benjamin, 2000) and larger predictability effects in aging (Choi et al., 2017). The underlying mechanisms for this effect are not fully understood, but one hypothesis is that older adults perceive their time as more limited than younger adults, and they consequently strive to spend more time with reviewing known information or positive content rather than gaining new information (i.e., theory of socioemotional selectivity, see Carstensen et al., 1999; Mather, 2004). For example, older adults have better recognition memory for positively,

[7] We investigated the possibility that older adults adopted a task adaptation strategy by investigating interactions between age, predictability, and trial number in the self-paced reading task. No interactions surfaced.

compared to negatively, valenced information (e.g., Charles et al., 2003; Fung & Carstensen, 2003). Such findings could suggest that aging reduces the willingness to engage in unknown experiences or negative emotions (i.e., "why get angry now?") and enhances the appreciation of known and positive aspects of life.

## Limitations

The older adults recruited for this study were younger compared to other old-adult samples in the extant literature (e.g., Fraundorf et al., 2019). They were also less disadvantaged in terms of reading speed as previous studies on old-adult populations have suggested (e.g., Kemper & Liu, 2007; Rayner et al., 2006). Even though we can probably exclude educational attainment as a possible driver of this effect (see Supplemental Material), it is possible that other cognitive variables in older adults, such as higher levels of print exposure or larger vocabulary levels, moderated the effect. It is a clear limitation of the present study that we could not account for possible influences of these other variables. The fact that the older adults recruited for this study participated from a website that promotes Psychology research clearly indicates high levels of intellectual curiosity in this sample. It remains to be seen if our current results generalize to other old-adult samples.

## Conclusion

When encoding sentences or text, readers create memory representations on multiple linguistic levels, for example, a surface level (i.e., word form) and a meaning representation level (e.g., Kintsch et al., 1990). The current findings add to our understanding of how aging impacts the memory representations that become created on the fly, how these representations are stored in long-term memory, and to what extent memory representations are influenced by different sentence and text processing strategies. While aging did not preferentially change integration costs during self-paced reading, older adults were more likely to encode gist-wise representations of text. This increased their memory for schema-congruent information, but it also increased false remembering. Nevertheless, rates of false remembering were attenuated in those aging readers who allocated more time to encode sentences on the fly. Together, our findings show that older adults represent text as a gist-wise approximation and that resource allocation during reading determines the accuracy of this representation. Future studies need to substantiate these findings with a more norm-typical sample of older adults.

## References


Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, *48*(1), 1–15. https://doi.org/10.1016/S0749-596X(02)00518-1

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Badham, S. P., Estes, Z., & Maylor, E. A. (2012). Integrative and semantic relations equally alleviate age-related associative memory deficits. *Psychology and Aging*, *27*(1), 141–152. https://doi.org/10.1037/a0023924

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., III, McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology*, *16*(3–5), 361–384. https://doi.org/10.1080/026432999380834

Balota, D. A., & Duchek, J. M. (1991). Semantic priming effects, lexical repetition effects, and contextual disambiguation effects in healthy aged individuals and individuals with senile dementia of the Alzheimer type. *Brain and Language*, *40*(2), 181–201. https://doi.org/10.1016/0093-934X(91)90124-J

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

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

Braver, T. S., & Barch, D. M. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience and Biobehavioral Reviews*, *26*(7), 809–817. https://doi.org/10.1016/S0149-7634(02)00067-2

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, *125*, Article 104334. https://doi.org/10.1016/j.jml.2022.104334

Brod, G., & Shing, Y. L. (2019). A boon and a bane: Comparing the effects of prior knowledge on memory across the lifespan. *Developmental Psychology*, *55*(6), 1326–1337. https://doi.org/10.1037/dev0000712

Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The influence of prior knowledge on memory: A developmental cognitive neuroscience perspective. *Frontiers in Behavioral Neuroscience*, *7*, Article 139. https://doi.org/10.3389/fnbeh.2013.00139

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), Article 9. https://doi.org/10.5334/joc.10

Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously. A theory of socioemotional selectivity. *American Psychologist*, *54*(3), 165–181. https://doi.org/10.1037/0003-066X.54.3.165

Chang, M., & Brainerd, C. J. (2021). Semantic and phonological false memory: A review of theory and data. *Journal of Memory and Language*, *119*, Article 104210. https://doi.org/10.1016/j.jml.2020.104210

Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, *132*(2), 310–324. https://doi.org/10.1037/0096-3445.132.2.310

Choi, W., Lowder, M. W., Ferreira, F., Swaab, T. Y., & Henderson, J. M. (2017). Effects of word predictability and preview lexicality on eye movements during reading: A comparison between young and older adults. *Psychology and Aging*, *32*(3), 232–242. https://doi.org/10.1037/pag0000160

Chung, Y. M. W., & Federmeier, K. D. (2023). Read carefully, because this is important! How value-driven strategies impact sentence memory. *Memory & Cognition*, *51*(7), 1511–1526. https://doi.org/10.3758/s13421-023-01409-3

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

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658–668. https://doi.org/10.1016/j.cognition.2006.10.010

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Dagerman, K. S., Macdonald, M. C., & Harm, M. W. (2006). Aging and the use of context in ambiguity resolution: Complex changes from simple slowing. *Cognitive Science*, *30*(2), 311–345. https://doi.org/10.1207/s15516709cog0000_46

Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, *5*(3), 305–312. https://doi.org/10.2466/pr0.1959.5.3.305

Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, *187*, 10–20. https://doi.org/10.1016/j.cognition.2019.01.001

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

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, *22*(1), 122–133. https://doi.org/10.1037/0882-7974.22.1.122

Failes, E., Sommers, M. S., & Jacoby, L. L. (2020). Blurring past and present: Using false memory to better understand false hearing in young and older adults. *Memory & Cognition*, *48*(8), 1403–1416. https://doi.org/10.3758/s13421-020-01068-8

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

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. https://doi.org/10.1111/1469-8986.3920133

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

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLOS ONE*, *8*(10), Article e77661. https://doi.org/10.1371/journal.pone.0077661

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). *LabVanced: A unified JavaScript framework for online studies* [Conference session]. International Conference on Computational Social Science (Cologne).

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15–52. https://doi.org/10.1016/j.cogpsych.2019.03.002

Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2019). Aging and recognition memory: A meta-analysis. *Psychological Bulletin*, *145*(4), 339–371. https://doi.org/10.1037/bul0000185

Fung, H. H., & Carstensen, L. L. (2003). Sending memorable messages to the old: Age differences in preferences and memory for advertisements. *Journal of Personality and Social Psychology*, *85*(1), 163–178. https://doi.org/10.1037/0022-3514.85.1.163

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*(7), 833–848. https://doi.org/10.3758/MC.38.7.833

Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language, and Hearing Research*, *40*(2), 423–431. https://doi.org/10.1044/jslhr.4002.423

Greene, N. R., Forsberg, A., Guitard, D., Naveh-Benjamin, M., & Cowan, N. (2024). A lifespan study of the confidence–accuracy relation in working memory and episodic long-term memory. *Journal of Experimental Psychology: General*, *153*(5), 1336–1360. https://doi.org/10.1037/xge0001551

Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, *148*(2), 325–341. https://doi.org/10.1037/xge0000498

Gunter, T. C., Jackson, J. L., & Mulder, G. (1992). An electrophysiological study of semantic processing in young and middle-aged academics. *Psychophysiology*, *29*(1), 38–54. https://doi.org/10.1111/j.1469-8986.1992.tb02009.x

Gunter, T. C., Jackson, J. L., & Mulder, G. (1995). Language, memory, and aging: An electrophysiological exploration of the N400 during reading of memory-demanding sentences. *Psychophysiology*, *32*(3), 215–229. https://doi.org/10.1111/j.1469-8986.1995.tb02951.x

Haeuser, K. I., & Kray, J. (2022a). Uninvited and unwanted: False memories for words predicted but not seen. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual meeting of the cognitive science society* (pp. 2401–2408). Cognitive Science Society. https://escholarship.org/uc/item/7w22b8gm

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

Haeuser, K. I., & Kray, J. (2021). Effects of prediction error on episodic memory retrieval: Evidence from sentence reading and word recognition. *Language, Cognition and Neuroscience*, *38*(4), 558–574. https://doi.org/10.1080/23273798.2021.1924387

Hartman, M., & Hasher, L. (1991). Aging and suppression: Memory for previously relevant information. *Psychology and Aging*, *6*(4), 587–594. https://doi.org/10.1037/0882-7974.6.4.587

Häuser, K. I., Demberg, V., & Kray, J. (2018). Surprisal modulates dual-task performance in older adults: Pupillometry shows age-related trade-offs in task performance and time-course of language processing. *Psychology and Aging*, *33*(8), 1168–1180. https://doi.org/10.1037/pag0000316

Heikkilä, J., Alho, K., Hyvönen, H., & Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Experimental Psychology*, *62*(2), 123–130. https://doi.org/10.1027/1618-3169/a000279

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–1326. https://doi.org/10.1002/hipo.20857

Höltje, G., Lubahn, B., & Mecklinger, A. (2019). The congruent, the incongruent, and the unexpected: Event-related potentials unveil the processes involved in schematic encoding. *Neuropsychologia*, *131*, 285–293. https://doi.org/10.1016/j.neuropsychologia.2019.05.013

Höltje, G., & Mecklinger, A. (2022). Benefits and costs of predictive processing: How sentential constraint and word expectedness affect memory formation. *Brain Research*, *1788*, Article 147942. https://doi.org/10.1016/j.brainres.2022.147942

Hubbard, R. J., Rommers, J., Jacobs, C. L., & Federmeier, K. D. (2019). Downstream behavioral and electrophysiological consequences of word prediction on recognition memory. *Frontiers in Human Neuroscience*, *13*, Article 291. https://doi.org/10.3389/fnhum.2019.00291

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

Hunt, R. R., & Lamb, C. A. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1359–1366. https://doi.org/10.1037/0278-7393.27.6.1359

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

Jongman, S. R., Copeland, A., Xu, Y., Payne, B. R., & Federmeier, K. D. (2023). Older adults show intraindividual variation in the use of predictive processing. *Experimental Aging Research*, *49*(5), 433–456. https://doi.org/10.1080/0361073X.2022.2137358

Kanske, P., & Kotz, S. A. (2010). Leipzig affective norms for German: A reliability study. *Behavior Research Methods*, *42*(4), 987–991. https://doi.org/10.3758/BRM.42.4.987

Kemper, S., & Liu, C.-J. (2007). Eye movements of young and older adults during reading. *Psychology and Aging*, *22*(1), 84–93. https://doi.org/10.1037/0882-7974.22.1.84

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*(2), 133–159. https://doi.org/10.1016/0749-596X(90)90069-C

Klever, L., Islam, J., Võ, M. L., & Billino, J. (2023). Aging attenuates the memory advantage for unexpected objects in real-world scenes. *Heliyon*, *9*(9), Article e20241. https://doi.org/10.1016/j.heliyon.2023.e20241

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*(4), 555–583. https://doi.org/10.1006/jmla.1997.2529

Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology and Aging*, *14*(2), 220–237. https://doi.org/10.1037/0882-7974.14.2.220

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

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

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

Light, L. L., LaVoie, D., Valencia-Laver, D., Owens, S. A., & Mead, G. (1992). Direct and indirect measures of memory for modality in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1284–1297. https://doi.org/10.1037/0278-7393.18.6.1284

Light, L. L., Valencia-Laver, D., & Zavis, D. (1991). Instantiation of general terms in young and older adults. *Psychology and Aging*, *6*(3), 337–351. https://doi.org/10.1037/0882-7974.6.3.337

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is Production: The missing link between language production and comprehension. *Scientific Reports*, *8*(1), Article 1079. https://doi.org/10.1038/s41598-018-19499-4

Mather, M. (2004). Aging and emotional memory. In D. Reisberg & P. Hertel (Eds.), *Memory and emotion* (pp. 272–307). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195158564.003.0009

Matzen, L. E., & Benjamin, A. S. (2013). Older and wiser: Older adults' episodic word memory benefits from sentence study contexts. *Psychology and Aging*, *28*(3), 754–767. https://doi.org/10.1037/a0032945

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *58*(1), 22–33. https://doi.org/10.1080/02724980443000151

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170

Ness, T., & Meltzer-Asscher, A. (2018). Lexical inhibition due to failed prediction: Behavioral evidence and ERP correlates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1269–1285. https://doi.org/10.1037/xlm0000525

Noh, S. R., & Stine-Morrow, E. A. (2009). Age differences in tracking characters during narrative comprehension. *Memory & Cognition*, *37*(6), 769–778. https://doi.org/10.3758/MC.37.6.769

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, *25*(6), 838–848. https://doi.org/10.3758/BF03211328

Packard, P. A., Rodríguez-Fornells, A., Bunzeck, N., Nicolás, B., de Diego-Balaguer, R., & Fuentemilla, L. (2017). Semantic congruence accelerates the onset of the neural signals of successful memory encoding. *The Journal of Neuroscience*, *37*(2), 291–301. https://doi.org/10.1523/JNEUROSCI.1622-16.2016

Payne, B. R., & Federmeier, K. D. (2017). Pace yourself: Intraindividual variability in context use revealed by self-paced event-related brain potentials. *Journal of Cognitive Neuroscience*, *29*(5), 837–854. https://doi.org/10.1162/jocn_a_01090

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. https://doi.org/10.1016/j.brainres.2018.02.021

Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., & Stine-Morrow, E. A. (2014). Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition*, *130*(2), 157–173. https://doi.org/10.1016/j.cognition.2013.10.005

Payne, B. R., & Silcox, J. W. (2019). Aging, context processing, and comprehension. In K. D. Federmeier (Ed.), *Psychology of learning and motivation* (Vol. 71, pp. 215–264). Academic Press. https://doi.org/10.1016/bs.plm.2019.07.001

Pichora-Fuller, K. M. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology*, *47*(Suppl. 2), S72–S82. https://doi.org/10.1080/14992020802307404

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*(1), 593–608. https://doi.org/10.1121/1.412282

Quent, J. A., Greve, A., & Henson, R. N. (2022). Shape of U: The nonmonotonic relationship between object–location memory and expectedness. *Psychological Science*, *33*(12), 2084–2097. https://doi.org/10.1177/09567976221109134

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–465. https://doi.org/10.1037/0882-7974.21.3.448

Rich, S., & Harris, J. (2021). Unexpected guests: When disconfirmed predictions linger. In T. Fitch, C. Lamm, H. Leder, & K. Teßmar-Raible (Eds.), *Proceedings of the 43rd annual meeting of the cognitive science* *society*. Cognitive Science Society. https://escholarship.org/uc/item/7q0000c8

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and Aging*, *27*(1), 33–45. https://doi.org/10.1037/a0026231

Rommers, J., & Federmeier, K. D. (2018a). Lingering expectations: A pseudo-repetition effect for words previously expected but not presented. *NeuroImage*, *183*, 263–272. https://doi.org/10.1016/j.neuroimage.2018.08.023

Rommers, J., & Federmeier, K. D. (2018b). Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. *Cortex*, *101*, 16–30. https://doi.org/10.1016/j.cortex.2017.12.018

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

Salthouse, T. A. (1992). Influence of processing speed on adult age differences in working memory. *Acta Psychologica*, *79*(2), 155–170. https://doi.org/10.1016/0001-6918(92)90030-H

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, *8*(4), 535–543. https://doi.org/10.1037/0894-4105.8.4.535

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*(1), 1–24. https://doi.org/10.1006/jmla.1998.2611

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

Shing, Y. L., Brod, G., & Greve, A. (2023). Prediction error and memory across the lifespan. *Neuroscience and Biobehavioral Reviews*, *155*, Article 105462. https://doi.org/10.1016/j.neubiorev.2023.105462

Skinner, D. J., & Price, J. (2019). The roles of meaningfulness and prior knowledge in younger and older adults' memory performance. *Applied Cognitive Psychology*, *33*(6), 1103–1112. https://doi.org/10.1002/acp.3552

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*(8), 311–327. https://doi.org/10.1111/lnc3.12151

Steen-Baker, A. A., Ng, S., Payne, B. R., Anderson, C. J., Federmeier, K. D., & Stine-Morrow, E. A. L. (2017). The effects of context on processing words during sentence reading among adults varying in age and literacy skill. *Psychology and Aging*, *32*(5), 460–472. https://doi.org/10.1037/pag0000184

Stine, E. A. L., & Hindman, J. (1994). Age differences in reading time allocation for propositionally dense sentences. *Aging, Neuropsychology and Cognition*, *1*(1), 2–16. https://doi.org/10.1080/09289919408251446

Stine-Morrow, E. A. L., Loveless, M. K., & Soederberg, L. M. (1996). Resource allocation in on-line reading by younger and older adults. *Psychology and Aging*, *11*(3), 475–486. https://doi.org/10.1037/0882-7974.11.3.475

Stine-Morrow, E. A. L., Shake, M. C., Miles, J. R., & Noh, S. R. (2006). Adult age differences in the effects of goals on self-regulated sentence processing. *Psychology and Aging*, *21*(4), 790–803. https://doi.org/10.1037/0882-7974.21.4.790

Stine-Morrow, E. A. L., Soederberg Miller, L. M., Gagne, D. D., & Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychology and Aging*, *23*(1), 131–153. https://doi.org/10.1037/0882-7974.23.1.131

Stine-Morrow, E. A. L., Soederberg Miller, L. M., & Nevin, J. A. (1999). The effects of context and feedback on age differences in spoken word recognition. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *54B*(2), P125–P134. https://doi.org/10.1093/geronb/54B.2.P125

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language*, *52*(1), 71–86. https://doi.org/10.1016/j.jml.2004.08.001

Tibon, R., Cooper, E., & Greve, A. (2017). Does semantic congruency accelerate episodic encoding, or increase semantic elaboration? *The Journal of Neuroscience*, *37*(19), 4861–4863. https://doi.org/10.1523/JNEUROSCI.0570-17.2017

Tun, P. A., & Wingfield, A. (1994). Speech recall under heavy load conditions: Age, predictability, and limits on dual-task interference. *Aging, Neuropsychology and Cognition*, *1*(1), 29–44. https://doi.org/10.1080/09289919408251448

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging*, *13*(2), 230–241. https://doi.org/10.1037/0882-7974.13.2.230

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

van de Meerendonk, N., Kolk, H. H., Vissers, C. T. W., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82. https://doi.org/10.1162/jocn.2008.21170

van Kesteren, M. T., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*(2), 272–292. https://doi.org/10.1037/0033-2909.120.2.272

Wingfield, A., Aberdeen, J. S., & Stine, E. A. (1991). Word onset gating and linguistic context in spoken word recognition by young and elderly adults. *Journal of Gerontology*, *46*(3), P127–P129. https://doi.org/10.1093/geronj/46.3.P127

Wingfield, A., Poon, L. W., Lombardi, L., & Lowe, D. (1985). Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time. *Journal of Gerontology*, *40*(5), 579–585. https://doi.org/10.1093/geronj/40.5.579

Wingfield, A., & Stine-Morrow, E. A. L. (2000). Language and speech. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 359–416). Erlbaum.

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Taylor & Francis. https://doi.org/10.4324/9781315165547

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research*, *41*(2), 105–128. https://doi.org/10.1007/s10936-011-9179-x

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975–988. https://doi.org/10.1037/a0029206

Wlotko, E. W., Lee, C. L., & Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Language and Linguistics Compass*, *4*(8), 623–638. https://doi.org/10.1111/j.1749-818X.2010.00224.x

Received October 17, 2023
Revision received July 12, 2024
Accepted July 12, 2024 ■