Metacognition and Learning (2024) 19:863–877
https://doi.org/10.1007/s11409-024-09391-0

Age-related differences in metacognitive reactivity in younger and older adults

Dillon H. Murphy[1] · **Matthew G. Rhodes**[2] · **Alan D. Castel**[1]

Received: 24 July 2023 / Accepted: 11 June 2024 / Published online: 1 July 2024

Abstract

When we monitor our learning, often measured via judgments of learning (JOLs), this metacognitive process can change what is remembered. For example, prior work has demonstrated that making JOLs enhances memory for related, but not unrelated, word pairs in younger adults. In the current study, we examined potential age-related differences in metacognitive reactivity. Younger and older adults studied lists of related and unrelated word pairs to remember for a later cued recall test where they would be presented with one of the words from the pair and be asked to recall its associate. Additionally, participants either made a JOL for each pair or had an inter-stimulus interval of equal duration as the JOL period. Results revealed that while making metacognitive judgments did not significantly affect memory in younger adults (i.e., no reactivity), this procedure impaired memory in older adults (i.e., negative reactivity), particularly for unrelated word pairs. Specifically, older adults demonstrated better cued recall when each word was followed by an inter-stimulus interval than when asked to predict the likelihood of remembering each word during the study phase. This may be a consequence of JOLs increasing task demands/cognitive load, which could reduce the elaborative encoding of associations between word pairs in older adults, but older adults' preserved or even enhanced semantic memory may mask negative reactivity for related word pairs. Future work is needed to better understand the mechanisms contributing to the reactivity effects in younger and older adults for different types of to-be-remembered information.

Keywords Reactivity · Metacognition · Monitoring · Aging · Semantic knowledge

✉ Dillon H. Murphy
Dillon.Murphy@asu.edu

1 Department of Psychology, University of California, Los Angeles, USA

2 Department of Psychology, Colorado State University, Fort Collins, USA

Age-related differences in metacognitive reactivity in younger and older adults

Whenever we predict whether we will remember something (and we often use those predictions to guide study decisions), we are relying on metacognition. Broadly, metacognition involves the awareness of our cognitive processes and abilities (Nelson & Narens, 1990; see also Dunlosky et al., 2016; Nelson, 1996; Rhodes, 2019) and can be considered in terms of *monitoring*, or the ability to predict whether something will be remembered later, and *control*, the self-regulation of learning that is typically based on monitoring assessments (Dunlosky & Ariel, 2011; Dunlosky et al., 2016; Metcalfe & Finn, 2008; Metcalfe, 2009; Nelson & Leonesio, 1988; Nelson et al., 1994; Robey et al., 2017; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999; Thiede et al., 2003; see Rhodes, 2016, for a review). While monitoring our learning and using this information to guide study decisions is an important component in maximizing memory utility, the monitoring process may inadvertently modify what we remember.

When attempting to remember information, metacognitive monitoring is frequently examined by soliciting predictions of future memory performance, often using judgments of learning (JOLs). These judgments are generally accurate such that learners can differentiate between what will be remembered and what will be forgotten on a memory test (see Koriat, 1997 for a discussion of the cues that inform learners' JOLs). However, the process of making these judgments can also influence memory, an effect known as metacognitive *reactivity* (e.g., Arbuckle & Cuddy, 1969; Double & Birney, 2019; Double et al., 2018; Janes et al., 2018; King et al., 1980; Maxwell & Huff, 2022a; Rivers et al., 2021; Soderstrom et al., 2015; Spellman & Bjork, 1992; Undorf et al., 2024; Witherby & Tauber, 2017). Specifically, when studying to-be-remembered information, if learners provide JOLs during encoding, they sometimes recall more information (positive reactivity) or less information (negative reactivity) compared to a group of learners not providing JOLs but who have equal exposure time to the information.

Although several prior studies have shown reactivity effects, other work has not observed memory benefits after making metacognitive judgments (e.g., Begg et al., 1992; Keleman & Weaver, 1997; Tauber & Rhodes, 2012). Further investigation has demonstrated that the presence or absence of metacognitive reactivity depends on the type of to-be-remembered information (but not whether a pure- or mixed-list design is employed; see Maxwell & Huff, 2022a). For example, Soderstrom et al. (2015) had participants study related (e.g., mother-father) and unrelated (twig-box) word pairs and assessed memory via a cued recall test whereby participants were given the cue (e.g., *twig*-) and asked to recall the target (e.g., *box*). Soderstrom et al. (2015) demonstrated that, relative to participants who did not make JOLs, participants providing JOLs showed enhanced memory for related pairs but not unrelated pairs (see Double et al., 2018 for similar results in a meta-analysis of eight different studies). Using a similar procedure, Mitchum et al. (2016) replicated the positive reactivity for related word pairs but also found negative reactivity (i.e., impaired recall) for unrelated word pairs. Thus, the relatedness of the to-be-remembered information can impact whether making JOLs enhances memory (see also Halamish & Undorf, 2023; Myers et al., 2020).

When studying a list of word pairs, the relationship between the two words is often a cue used to inform JOLs (Dunlosky & Matvey, 2001; Mueller et al., 2013) and the degree to which the words are related typically positively predicts memory (e.g., Soderstrom et al., 2015). To account for differences in reactivity for different types of materials, Soderstrom et al. (2015) suggested that learners use the relatedness of word pairs to inform their JOLs, strengthening the relationship and subsequent encoding for only related pairs (see also Halamish & Undorf, 2023). If JOLs strengthen the relationship between related pairs, this may make a word's associate more accessible during retrieval, leading to positive reactivity (but reactivity may not be present when examining memory using free-recall paradigms, e.g., Myers et al., 2020; Tauber & Rhodes, 2012; but see Begg et al., 1989). In contrast, because there is no semantic connection for unrelated pairs, JOLs do not have a relationship to strengthen. Thus, JOLs strengthening the relationship between related word pairs is a potential mechanism of reactivity.

Related to the idea that cue-strengthening can produce reactivity, Maxwell and Huff (2022b) explored whether JOL reactivity patterns are unique to memory forecasting or if they could be attributed to relational encoding strategies applied to both related and unrelated word pairs. Specifically, Maxwell and Huff (2022b) compared JOLs not only with a control group but also with groups engaged in different relational encoding (i.e., judgments of associative memory, frequency judgments; see also Halamish & Undorf, 2023). Results revealed that JOLs did not uniquely influence recall through memory forecasting. Rather, JOL reactivity can also occur in tasks involving relational encoding, suggesting that memory reactivity could partly reflect the cognitive processes involved in evaluating relationships within material, rather than solely the act of predicting future recall.

Together, prior work suggests that metacognitive reactivity likely reflects multiple mechanisms (see Janes et al., 2018; Myers et al., 2020). Specifically, in addition to the cue strengthening account, other theories of reactivity have suggested that JOLs benefit recall as a result of "metacognition modifying attention" (Castel et al., 2012), from enhancing the encoding of item-specific information (Senkova & Otani, 2021; Li et al., 2024b), changing learning goals (i.e., shifting attention to information that is viewed as easy or moderately difficult to master at the cost of more challenging items; Mitchum et al., 2016; but se Li et al., 2024a), and/or by updating agendas and changing how learners prioritize items (i.e., they focus on information consistent with their agenda, e.g., Ariel et al., 2009; Murphy & Castel, 2021; Murphy et al., 2023). Another potential mechanism of reactivity may be that making JOLs increases learners' cognitive load, resulting in dual-task costs (see Janes et al., 2018), leading to negative reactivity. Again, each of the mechanisms has received support, suggesting multiple mechanisms underlying reactivity, but there may be age-related differences in metacognitive reactivity.

Prior work examining JOLs indicates that younger and older adults are similarly capable of monitoring their learning such that discrimination, or the relative accuracy of memory judgments (i.e., identifying which items will be remembered and which will be forgotten), is preserved in older adults (Connor et al., 1997; Hertzog et al., 2010). Given their similar monitoring abilities, it may be the case that both younger and older adults show similar levels of metacognitive reactivity. However, because making metacognitive judgments can enhance memory for related word pairs in younger adults (Mitchum et al., 2016; Soderstrom et al., 2015), positive reactivity could be more pronounced in older adults due to age-related differences in semantic memory.

Older adults often experience episodic memory deficits (see Balota et al., 2000; Craik, 2020) but semantic memory tends to be preserved or even enhanced in older adults (Dixon, 2003; Park, 2000; Salthouse, 2004; Spreng & Turner, 2019; Staudinger et al., 1989), and older adults can rely on their wealth of prior knowledge more than younger adults (Umanath & Marsh, 2014). Thus, older adults' semantic memory abilities may lead to greater strengthening of the relationship between related word pairs, especially when making JOLs. Specifically, if reactivity arises because JOLs strengthen the relationship between related word pairs (one of the mechanisms that has gained the most support, e.g., Soderstrom et al., 2015), older adults' semantic memory abilities may result in more reactivity if making JOLs emphasizes the connection between word pairs more than it does for younger adults.

When older adults use semantic cue-target relationships to make JOLs, this process might enhance these connections more effectively than in younger adults, either through direct cue-strengthening or within the encoded memory trace (relational encoding; cf. Maxwell & Huff, 2022b). This enhancement likely depends on the initial strength of the semantic relationship, which is potentially more robust in older adults, aligning with the notion that JOL-induced strengthening necessitates an underlying semantic link to be effective—explaining the lack of reactivity for unrelated pairs. Yet, this effect may follow a non-linear pattern: a basic level of association might be necessary to highlight the relationship during JOLs, but very strong pre-existing connections might not benefit as much, or could even detract from, the additional cue strengthening prompted by JOLs.

Older adults' reliance on existing semantic knowledge to learn related word pairs could also enhance their episodic memory, which is usually diminished relative to younger adults (e.g., Naveh-Benjamin, 2000). Since older adults often depend more on semantic memory to support their poorer episodic memory, the potential memory boost afforded by reinforcing cue-target connections when making JOLs may offer greater retrieval advantages for older adults despite similar levels of cue-target strengthening observed in both age groups. In contrast, when learning unrelated pairs, there is minimal schematic support so older adults may show no reactivity or even negative reactivity for unrelated pairs, similar to younger adults.

Despite JOLs potentially strengthening the cue-target relationship, making JOLs increases the cognitive load of the learner (making JOLs is a secondary task in addition to encoding the presented information), potentially yielding poorer encoding and subsequent retrieval of the to-be-remembered information. Specifically, if making JOLs increases cognitive load, leading to dual-task detriments (see Janes et al., 2018), older adults might exhibit diminished or even negative reactivity compared with younger adults due to older adults' reduced cognitive abilities (see Hess, 2005; Park & Festini, 2017; Salthouse, 2010, 2019; Thomas & Gutchess, 2020). Thus, making JOLs could result in negative reactivity, particularly in older adults, for both related and unrelated word pairs due to increased cognitive load.

Some prior work has examined the effects of JOLs on memory performance across younger and older adults. Specifically, Tauber and Witherby (2019) presented younger and older adults with lists of word pairs, tasking them with making predictions about their likelihood of remembering these pairs for a later cued recall test. Pairs were displayed for 10 s each, with participants not making JOLs studying the pair for the entire duration while those in the JOL group made their judgments after an initial 5-second study period, with the word pair still visible for another 5 s while participants provided their JOLs. Results revealed that while JOLs improved recall for younger adults, making JOLs did not modify older adults' memory performance, suggesting age-related differences in reactivity. However, the to-be-remembered information in Tauber and Witherby's (2019) study was primarily highly related word pairs, necessitating further work to examine potential age-related differences in reactivity for unrelated word pairs. Additionally, having JOLs occur during word pair presentation may be problematic because one group studied each pair for 10 s while the other group studied each pair for 5 s before being asked to make a JOL (potentially returning to study the word pair after providing the JOL, but that study time is still lost), meaning the time available to focus on encoding each pair was less than 10 s for the JOL group. A design equating time where the word pairs are on-screen with no secondary task to complete may yield different patterns of reactivity.

The current study

In the current study, we aimed to further examine how reactivity differs between younger and older adults. We presented younger and older adults with three lists of 10 related and 10 unrelated word pairs to remember for a later cued recall test whereby they would be presented with the cue word from the pair and asked to recall its associate. Participants studied each pair for 5 s and then were either given 7 s to predict the likelihood of later remembering the pair (i.e., a JOL) or a 7-second inter-stimulus interval (i.e., without the pair displayed) to equate study time across conditions. Consistent with prior work (e.g., Myers et al., 2020; Soderstrom et al., 2015), we expected JOLs to enhance memory for related but not unrelated pairs relative to the no-JOL group. Regarding potential age-related differences, with the cognitive load during encoding potentially reduced relative to Tauber and Witherby (2019)— rather than 60 word pairs per list, the present study used 20—we expected that older adults may demonstrate more reactivity as a product of making JOLs relative to younger adults, particularly for related pairs since relatedness might activate older adults' semantic memory. Moreover, given that older adults' semantic memory is preserved or enhanced (Dixon, 2003; Park, 2000; Salthouse, 2004; Spreng & Turner, 2019; Staudinger et al., 1989), we anticipated that older adults may use relatedness as a cue more than younger adults, potentially leading to greater reactivity for related word pairs (but possibly negative reactivity for unrelated pairs).

Method

Participants

Our design was pre-registered on OSF. A pilot study with a similar procedure is also available on OSF. Data were collected in the Fall of 2022. Younger adults were 256 undergraduate students (M_{age} = 19.99, SD_{age} = 1.78, range: 18–31; 204 female, 52 male; 1 American Indian/Alaskan Native, 98 Asian/Pacific Islander, 10 Black, 53 Hispanic, 74 white, 20 other/unknown; in terms of the highest level of education achieved, 1 Some High School, 47 High School Graduate, 147 some college but no degree, 46 Associates degree, 15 Bachelor's degree) recruited from our university's Human Subjects Pool. Participants were tested online and received course credit for their participation. Older adults (n = 225; M_{age} = 72.90, SD_{age} = 5.60, range: 61–96; 135 female, 88 male, 2 other; 2 American Indian/Alaskan Native, 1 Asian/Pacific Islander, 5 Black, 215 white, 2 other/unknown; 5 Some High School, 33 High School Graduate, 58 some college but no degree, 34 Associates degree, 58 Bachelor's degree, 37 Graduate degree) were recruited from Amazon's Cloud Research (Chandler et al., 2019), a website that allows users to complete small tasks for pay.

Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This process resulted in 24 exclusions (five younger adults, 19 older adults). We also excluded participants who provided very few JOLs (less than 50% of words) throughout the task, which resulted in two exclusions (one younger adult and one older adult). We also excluded five participants who did not recall any words throughout the task (one younger adult and four older adults). Finally, four participants (two younger adults and two older adults) were excluded for failing to pass an attention check question (described below). As per our preregistration, we aimed to collect around 125 participants per condition. We may have been underpowered to detect the 3-way interaction in our pilot study so we want to ensure a well-powered design and account for the possibility of having to exclude several participants given the new exclusion criteria.

Materials

Stimuli were taken from Myers et al. (2020). Word pairs consisted of 30 related and 30 unrelated word pairs similar in average forward association, frequency, concreteness, and target length. Words were randomly chosen for each list and the order of items was randomized.

Procedure

Participants were presented with lists of 20 word pairs to remember for a later test whereby participants would be shown the left word from each pair and asked to recall the associated word. Each list contained 10 related and 10 unrelated word pairs and each pair was presented for 5 s. The order of word pairs within each list was randomized. Following the presentation of each pair, some participants predicted the likelihood of remembering the pair on the test (i.e., a JOL; $n_{young} = 128$, $n_{old} = 119$). Specifically, participants were asked "How likely are you to remember this pair?". These judgments were made on a scale of 0 (not at all likely) to 100 (very likely). Participants were given 7 s to provide their JOL; the word pair did not remain on the screen while participants provided their judgments. Rather than making a JOL after each word, another group of participants ($n_{young} = 128$, $n_{old} = 106$) had a 7-second inter-stimulus interval to match the length of the study phase for each group.

After studying the 20 word pairs, all participants completed a 30-second distraction task requiring them to rearrange the digits of several three-digit numbers in descending order (e.g., 123 would be rearranged to 321). Participants were given 3 s to view each of the 10, three-digit numbers and to subsequently rearrange the digits. Following the distractor task, all participants completed the cued recall test. On the test, participants were shown the left word from each pair (i.e., the cue word), in random order, and were asked to recall its associate (i.e., the target word); participants were given as much time as they needed to recall each pair. This was repeated for a total of three study-test cycles.

At the end of the task, participants completed the Dementia Worry Scale (Roberts & Maxfield, 2021) where participants rate (on a 1–5 Likert Scale) 12 statements regarding Alzheimer's disease and related dementias in terms of how typical they are of them (this questionnaire was administered to address a research question unrelated to the current study and thus is not reported, but those data are available on OSF). We also included an attention check within the Dementia Worry Scale. Specifically, participants were asked to select the answer to this question: "1 + 1 = ?"; participants selected from the following answer choices: 1, 2, 3, 4, or 5. If participants answered incorrectly, they were excluded from analysis.

Results

Target recall as a function of age and relatedness for the JOL and no-JOL groups is shown in Fig. 1. Data were analyzed using JASP (Love et al., 2019) and all information needed to reproduce the analyses is available on OSF. Mean JOLs for related and unrelated pairs as well as Gamma correlations between JOLs and recall (a measure of resolution, or the relative accuracy of participants' predictions) are shown in the Appendix.

Analytical plan

For transparency, we have chosen to report both frequentist and Bayesian results, providing a more comprehensive view of the data. Specifically, to examine the strength of the evidence for each effect in our inferential tests, we also computed the Bayes Factor (a ratio of the marginal likelihood of the null model and a model suggesting group differences) compared to a null model using JASP. We provide BF_{01} when inferential statistics favor the null hypothesis (which would be supported by a large BF_{01}) and BF_{10} when inferential statistics favor the alternative hypothesis (which would be supported by a large BF_{10}; for more information on interpreting Bayes factors, see Kass & Raftery, 1995). Generally, BF_{01} or BF_{10} values of less than 3 are considered anecdotal evidence while values greater than 3 indicate substantial evidence for said effect.

Beyond frequentist and Bayesian statistics, we are also interested in the practical significance of effects which involves evaluating the size and importance of an effect in a real-world context, beyond just determining its statistical significance (see Makowski et al., 2019). Thus, we report confidence intervals for effect sizes as this can provide insight

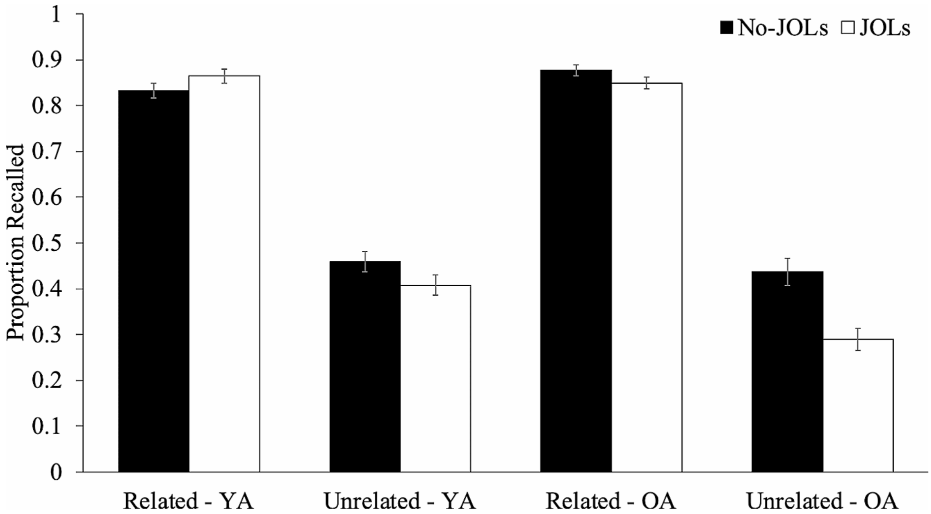

Fig. 1 Cued recall as a function of age (YA=younger adults, OA=older adults) and relatedness for the JOL and no-JOL groups. Error bars reflect the standard error of the mean

into the precision and reliability of the estimated effect—a 95% confidence interval ($CI_{95\%}$) for an effect size provides an estimated range of values for the effect size that is believed to cover the true effect size in 95% of similar experiments. In addition to confidence intervals, we also include credible intervals ($CrI_{95\%}$) for each effect size (Tendeiro et al., 2024 suggest reporting the size of each effect, confidence and/or credible intervals, as well as the Bayes factor). In Bayesian statistics, a credible interval is a range of values in which an unknown parameter falls with a certain probability, reflecting the degree of certainty or "credibility" we have about the parameter's true value. Unlike confidence intervals in frequentist statistics, which are about repeated sampling, a credible interval directly provides the probability of the true effect size lying within the given interval.

We recognize that *p*-values, Bayes Factor, confidence intervals, and credible intervals offer different types of evidence. However, to anticipate, there are instances in our ANOVA where the *p*-value indicates that the observed data are unlikely under the null hypothesis while Bayes Factor suggests that the evidence is not strong enough to decisively favor one hypothesis over the other. This discrepancy highlights the importance of considering both statistical significance and the strength of evidence when making inferences about hypotheses. In these instances, we suggest a cautious interpretation of the results. We also note that calculating Bayes Factor can be complicated when more than two factors are added to the model, and reasonable researchers may suggest different approaches.

Model terms

We conducted a 2 (relatedness: unrelated, related) x 2 (JOL versus no-JOL) x 2 (age: young, old) mixed-factor ANOVA. For each main effect, to further compare the levels of each variable, we supplemented the results of the ANOVA with a separate *t*-test which allowed us to compute a confidence interval for the effect size (Cohen's *d*) as well as credible intervals (JASP provides software to easily compute these statistics for *t*-tests but not ANOVA). Similarly, for each interaction, we conducted a separate *t*-test for each simple effect comparison. Here, the extra *t*-tests directly analyze the differences between paired observations, potentially providing a more straightforward measure of the effect size between two conditions while the repeated measures ANOVA considers the variance within each condition and between subjects, which might dilute effects. We acknowledge that conducting these separate *t*-tests increases our Type I error rate, but we note that we rely on them primarily to support the results displayed by the ANOVA. We include these supplemental tests because they provide important additional evidence for each result, particularly when there are instances where the *p*-value and Bayes Factor from the ANOVA do not yield an identical interpretation.

Main effects

Results revealed that related word pairs were better recalled ($M=0.86$, $SD=0.16$) than unrelated word pairs ($M=0.40$, $SD=0.27$), [$F(1, 477)=2{,}182.68, p<.001, \eta_p^2=0.82, BF_{10}>100$]; a separate *t*-test corroborated this effect [$t(480)=44.73$, $p<.001$, $d=2.03$, $CI_{95\%}d=1.88$–2.20, $BF_{10}>100$, $CrI_{95\%}=1.88$–2.19]. Younger adults recalled a similar proportion of word pairs ($M=0.64$, $SD=0.19$) as older adults ($M=0.61$, $SD=0.19$), [$F(1, 477)=2.68$, $p=.103$, $\eta_p^2=0.01$, $BF_{01}=4.79$]; a separate *t*-test generally corroborated this effect [$t(479) = -1.77$,

$p=.078$, $d=0.16$, $CI_{95\%}d=-0.34-0.02$, $BF_{01}=2.17$, $CrI_{95\%}=-0.33-0.02$], though the evidence is mixed—it appears that there could be a small recall advantage for younger adults (which would be consistent with prior work; see Naveh-Benjamin, 2000). Participants making JOLs recalled fewer word pairs ($M=0.60$, $SD=0.18$) than participants not making JOLs ($M=0.65$, $SD=0.19$), [$F(1, 477)=8.20$, $p=.004$, $\eta_p^2=0.02$, $BF_{10}=0.88$], exemplifying negative reactivity; a separate t-test corroborated this effect [$t(479)=2.75$, $p=.006$, $d=0.25$, $CI_{95\%}d=0.07-0.43$, $BF_{10}=3.88$, $CrI_{95\%}=0.07-0.42$].

Interactions

Whether or not participants made JOLs interacted with age group [$F(1, 477)=5.31$, $p=.022$, $\eta_p^2=0.01$, $BF_{10}=0.60$] such that when making JOLs, older adults' recall was impaired relative to the no JOL group (negative reactivity), [$p_{holm}=0.003$, $d=0.41$]; a separate t-test corroborated this effect [$t(223)=3.62$, $p<.001$, $d=0.48$, $CI_{95\%}d=0.22-0.75$, $BF_{10}=61.49$, $CrI_{95\%}=0.20-0.73$]. However, younger adults did not show reactivity [$p_{holm}>0.999$, $d=0.04$]; a separate t-test corroborated this effect [$t(254)=0.40$, $p=.689$, $d=0.05$, $CI_{95\%}d=-0.20-0.30$, $BF_{01}=6.76$, $CrI_{95\%}=-0.19-0.29$].

Relatedness interacted with age group [$F(1, 477)=18.71$, $p<.001$, $\eta_p^2=0.04$, $BF_{10}>100$] such that older adults recalled fewer unrelated pairs than younger adults [$p_{holm}<0.001$, $d=0.33$]; a separate t-test corroborated this effect [$t(479)=-3.07$, $p=.002$, $d=0.28$, $CI_{95\%}d=-0.46--0.10$, $BF_{10}=9.54$, $CrI_{95\%}=-0.45--0.10$]. However, recall was similar between age groups for related pairs [$p_{holm}=0.462$, $d=0.07$]; a separate t-test corroborated this effect [$t(479)=0.95$, $p=.341$, $d=0.09$, $CI_{95\%}d=-0.09-0.27$, $BF_{01}=6.35$, $CrI_{95\%}=-0.09-0.26$].

Next, whether or not participants made JOLs interacted with relatedness [$F(1, 477)=26.64$, $p<.001$, $\eta_p^2=0.05$, $BF_{10}>100$] such that when participants made JOLs, there was negative reactivity for unrelated word pairs [$p_{holm}<0.001$, $d=0.46$]; a separate t-test corroborated this effect [$t(479)=4.08$, $p<.001$, $d=0.37$, $CI_{95\%}d=0.19-0.55$, $BF_{10}>100$, $CrI_{95\%}=0.19-0.54$]. However, there was no reactivity for related word pairs [$p_{holm}=0.929$, $d=0.01$]; a separate t-test corroborated this effect [$t(479)=-0.29$, $p=.775$, $d=-0.03$, $CI_{95\%}d=-0.21-0.15$, $BF_{01}=9.49$, $CrI_{95\%}=-0.20-0.15$].

There was not a significant three-way interaction between relatedness, age, and whether or not participants made JOLs [$F(1, 477)=0.94$, $p=.332$, $\eta_p^2<0.01$, $BF_{01}=4.00$]. However, given our hypotheses and visual examination of Fig. 1, as an exploratory analysis, we conducted post hoc tests (as well as the separate t-tests) to examine reactivity as a function of word relatedness for younger and older adults. For older adults, making JOLs did not impact recall for related words [$p_{holm}>0.999$, $d=0.13$]; a separate t-test corroborated this effect [$t(223)=1.56$, $p=.120$, $d=0.21$, $CI_{95\%}d=-0.05-0.47$, $BF_{01}=2.18$, $CrI_{95\%}=-0.06-0.45$]. However, making JOLs impaired recall for unrelated words in older adults [$p_{holm}<0.001$, $d=0.69$]; a separate t-test corroborated this effect [$t(223)=4.01$, $p<.001$, $d=0.54$, $CI_{95\%}d=0.27-0.80$, $BF_{10}>100$, $CrI_{95\%}=0.25-0.78$]. For younger adults, making JOLs did not lead to reactivity for related words [$p_{holm}>0.999$, $d=-0.15$]; a separate t-test corroborated this effect [$t(254)=-1.45$, $p=.148$, $d=-0.18$, $CI_{95\%}d=-0.43-0.06$, $BF_{01}=2.70$, $CrI_{95\%}=-0.41-0.07$]. Similarly, making JOLs did not lead to reactivity for unrelated words in younger adults [$p_{holm}=0.543$, $d=0.24$]; a separate t-test corroborated this effect [$t(254)=1.65$, $p=.101$, $d=0.21$, $CI_{95\%}d=-0.04-0.45$, $BF_{01}=2.03$, $CrI_{95\%}=-0.04-0.44$].

We suggest caution when interpreting these results as this analysis of the three-way interaction was exploratory, although we believed it was worthwhile given the theoretical justification for expecting effects to vary across subgroups. We also note that the non-significant three-way interaction could be due to a lack of statistical power rather than the absence of a real effect. Specifically, although we had a large sample size ($n=481$), three-way interactions can be difficult to detect—detecting higher-order interactions generally requires more statistical power than detecting main effects or lower-order interactions.

Given this potential power issue, we conducted a power analysis to determine what the recommended sample size would have been to detect a small effect. For a small effect size (Cohen's $d=0.20$) in a 2 (age; between-subjects) x 2 (JOL condition; between-subjects) x 2 (relatedness; within-subjects) design, aiming for a significance level of 0.05 and a desired power of 80%, we would have needed approximately 361 participants in each of the four between-subjects conditions (1,442 participants total). Given that we had 481 participants, the power we obtained is approximately 30%. Thus, detecting a small three-way interaction in the present design requires a massive sample size and we may not have achieved sufficient power to detect a small three-way interaction in the present design. We also note that certain elements of our design may have also reduced power. For example, there may be a ceiling effect in the recall of related word pairs (86%), potentially reducing the sensitivity of that measure. Future work should aim to replicate these effects using a design with more statistical power.

Discussion

Overall, we found negative reactivity, and there was some evidence that this negative reactivity was specific to unrelated word pairs (consistent with some prior work, see Mitchum et al., 2016) with no reactivity for related word pairs. However, we note that while there have been instances of negative reactivity (e.g., Rivers et al., 2021), the predominant trend observed is one of positive reactivity towards related word pairs and a neutral response to unrelated pairs, as highlighted in a recent meta-analysis (Double et al., 2018).

Regarding age-related differences, while making metacognitive judgments did not significantly impact memory in younger adults, this procedure impaired memory in older adults. Specifically, older adults demonstrated better cued recall when each word was followed by an unfilled inter-stimulus interval than when they predicted the likelihood of remembering each word during the study phase. This contrasts with the previous work of Tauber and Witherby (2019) who found that, while younger adults making JOLs recalled significantly more than younger adults not making JOLs, older adults' memory performance was not impacted by making JOLs. This discrepancy may highlight the variability in how JOLs influence memory across different contexts (e.g., in the lab, online testing, using different paired associates/items, self-paced study time, use of feedback, multiple study-test cycles, etc.) or suggest that the cognitive mechanisms underlying reactivity effects may be more complex, particularly in diverse older populations and samples that could vary in terms of age range, level of education, vocabulary, and concerns about memory. Additionally, reactivity may be differentially impacted by stereotype threat when asked to make judgments about memory performance (Fourquet et al., 2020), potentially creating individual differences that could be examined in future research.

The present study introduced notable procedural variations compared to previous research on reactivity (e.g., Soderstrom et al., 2015; Tauber & Witherby, 2019). For example, we used shorter lists of word pairs and in our no-JOL condition, participants had an unfilled interval rather than the word-pair remaining on-screen, and the retention interval in our study was significantly shorter, set at 30 s, as opposed to longer intervals (e.g., 3 to 5 min). These procedural differences may have influenced the results, potentially accounting for discrepancies between our findings and those of earlier work. Specifically, the reduced list size may have decreased cognitive load (or led to a ceiling effect for related word pairs), potentially creating an environment where memory performance is elevated, leaving little room for positive reactivity and potentially making negative reactivity more likely. Moreover, the absence of additional learning opportunities in the no-JOL condition could have reduced recall for that group, and the shorter retention interval may have affected memory consolidation processes. It is also important to highlight that Tauber and Witherby (2019) predominantly used strongly related word pairs in their study. While we present a similar trend of older adults not showing reactivity for related pairs, we also show some evidence that older adults display negative reactivity for unrelated pairs. Together, these factors suggest that variations in study design may not only contribute to the observed differences in results but also underscore the complexity of reactivity effects across different experimental conditions.

The age-related differences observed in the present study appear to challenge explanations of JOL reactivity that emphasize the "cue-strengthening" account (i.e., although the pattern of data for younger adults is what would somewhat be anticipated by this account, older adults displayed some negative reactivity). Specifically, Soderstrom et al. (2015) suggested that the process of making JOLs draws learners' attention to the relationship between the word pairs (as this is a cue that informs JOLs, see Koriat, 1997) and this subsequently enhances memory for related word pairs (which have a semantic connection). In the present study, rather than strengthening the cues, there may be an interplay between cognitive load and metacognitive processes that led to negative reactivity. Again, older adults experience increased cognitive load during certain tasks due to age-related declines in working memory capacity and processing speed (Balota et al., 2000; Craik, 2020). As such, when older adults made JOLs during the study phase, this additional cognitive demand might have interfered with their ability to engage in elaborative encoding processes, particularly for unrelated word pairs. Consequently, this reduced encoding effort could have led to negative reactivity, impairing memory but preserving metacognitive accuracy (as shown in the Appendix).

One possibility is that making metacognitive predictions of memory performance increases cognitive load and reduces elaborative processing of the associations between word pairs for older adults. In contrast, when each word pair is followed by an inter-stimulus interval, older adults may be able to use that time to rehearse and elaborate on the associations between the items, even when they are unrelated. Given that older adults have fewer cognitive resources available during time-based encoding tasks relative to younger adults (Balota et al., 2000; Craik, 2020), JOLs may detract from elaborative encoding operations resulting in poorer memory performance. Thus, there may be a trade-off between aging and cognitive load such that younger adults can manage the JOL task while also learning the words, but the two tasks are more attentionally demanding for older adults, resulting in negative reactivity.

Although we present evidence that older adults displayed negative reactivity and younger adults did not display reactivity, the present data is unable to discern whether these age-related differences in reactivity were the result of older adults facing greater costs or deriving fewer benefits from making JOLs compared to younger adults. Future work could examine cognitive load during encoding to explore whether the process of making metacognitive judgments increases cognitive load, potentially reducing recall. For example, future work using divided attention tasks (see Naveh-Benjamin et al., 2005) may elucidate the attentional load explanation of negative reactivity for older adults.

Prior work suggests that making JOLs can selectively improve memory for related word pairs (e.g., Halamish & Undorf, 2023; Soderstrom et al., 2015), although this was not the case in the present study. However, the possibility of a ceiling effect on the relatedness of word pairs should be considered. For individuals who have formed strong and enduring associations between certain word pairs over many decades, additional exposures to these highly related pairs may not lead to a significant increase in their relatedness. Consequently, the potential for positive reactivity, which is often observed when related word pairs are enhanced through metacognitive judgments, may be limited for older adults who already possess highly interconnected semantic networks (i.e., older adults' existing semantic networks already facilitate effective recall without the need for additional JOL-induced elaboration). Relatedly, younger adults' generally superior memory abilities relative to younger adults (Naveh-Benjamin, 2000) may have led to a ceiling effect in the recall of related word pairs (85%), potentially creating an environment with little opportunity for positive reactivity (i.e., memory is already at ceiling). Future work should examine the potential reactive effects of JOLs using procedures yielding lower performance for related word pairs in both younger and older adults.

In sum, the present study demonstrated that while making metacognitive judgments did not impact memory in younger adults (we found neither positive nor negative reactivity), engaging in this metacognitive process can impair certain forms of memory in older adults, and this negative reactivity in older adults may be specific to unrelated word pairs. This may arise as a consequence of JOLs increasing task demands/cognitive load, which could reduce the encoding of word pairs (though older adults' superior semantic memory may mask negative reactivity for related word pairs). Specifically, when making JOLs, older adults may not be able to engage in elaborative associative memory processes that link unrelated units of information which could lead to negative reactivity. However, future work is needed to better understand the mechanisms contributing to reactivity effects in younger and older adults for different types of to-be-remembered information.

Appendix

Means and standard deviations (in parentheses) for JOLs for related and unrelated word pairs as well as Gamma correlations between JOLs and recall. There were no significant differences in Gammas (a measure of the relative accuracy of JOLs) between younger and older adults [$t(242)=1.44$, $p=.152$, $d=0.18$, $CI_{95\%}d=-0.07 - 0.44$, $BF_{01}=2.69$, $CrI_{95\%} = -0.07 - 0.42$].

	JOLs Related Pairs	JOLs Unrelated Pairs	Gamma
Younger adults	71.99 (17.73)	34.94 (17.04)	0.65 (0.22)
Older adults	70.57 (28.21)	28.56 (24.72)	0.70 (0.33)

Declarations

Ethics approval and consent to participate This research was supported in part by the National Institutes of Health (National Institute on Aging; Award Number R01 AG044335 to Alan D. Castel). The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript. Informed consent was acquired, and the study was completed in accordance with the UCLA Institutional Review Board.

Open practices statement The experiment reported in this article was formally preregistered, and the stimuli and data have been made available on the Open Science Framework here.

References

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*, 126–131.

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, *138*, 432–447.

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy young and older adults. *The Oxford Handbook of Memory*, 395–410.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*, 610–632.

Begg, I. M., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology*, *4*, 195–218.

Castel, A. D., McGillivray, S., & Friedman, M. C. (2012). Metamemory and memory efficiency in older adults: Learning about the benefits of priority processing and value-directed remembering. In M. Naveh-Benjamin, & N. Ohta (Eds.), *Memory and aging: Current issues and future directions* (pp. 245–270). Psychology.

Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond mechanical Turk. *Behavior Research Methods*, *51*, 2022–2038.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, *12*, 50–71.

Craik, F. I. (2020). Remembering: An activity of mind and brain. *Annual Review of Psychology*, *71*, 1–24.

Dixon, R. A. (2003). Themes in the aging of intelligence: Robust decline with intriguing possibilities. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 151–167). American Psychological Association.

Double, K. S., & Birney, D. P. (2019). Reactivity to measures of metacognition. *Frontiers in Psychology*, *10*, 2755.

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgments of learning. *Memory (Hove, England)*, *26*, 741–750.

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation: Vol. 54. The psychology of learning and motivation: Advances in research and theory* (pp. 103–140). Elsevier Academic.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*, 1180–1191.

Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: Problems and pitfalls. In J. Dunlosky, & S. K. Tauber (Eds.), *Oxford library of psychology. The Oxford handbook of metamemory* (pp. 23–37). Oxford University Press.

Fourquet, N. Y., Patterson, T. K., Li, C., Castel, A. D., & Knowlton, B. J. (2020). Effects of age-related stereotype threat on metacognition. *Frontiers in Psychology*, *11*, 604978.

Halamish, V., & Undorf, M. (2023). Why do judgments of learning modify memory? Evidence from identical pairs and relatedness judgments. *Journal of Experimental Psychology: Learning Memory and Cognition*, *49*, 547–556.

Hertzog, C., Sinclair, S. M., & Dunlosky, J. (2010). Age differences in the monitoring of learning: Cross-sectional evidence of spared resolution across the adult life span. *Developmental Psychology*, *46*, 939–948.

Hess, T. M. (2005). Memory and aging in context. *Psychological Bulletin*, *131*, 383–406.

Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review*, *25*, 2356–2364.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Keleman, W. L., & Weaver, C. A. (1997). Enhanced metamemory at delays: Why do judgments of learning improve over time? *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*, 1394–1409.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, *93*, 329–343.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370.

Li, B., Shanks, D. R., Zhao, W., Hu, X., Luo, L., & Yang, C. (2024a). Do changed learning goals explain why metamemory judgments reactively affect memory? *Journal of Memory and Language*, *136*, 104506.

Li, B., Pastötter, B., Zhong, Y., Su, N., Huang, T., Zhao, W., & Yang, C. (2024b). Judgments of learning reactively improve memory by enhancing Learning Engagement and Inducing Elaborative Processing: Evidence from an EEG study. *Journal of Intelligence*, *12*, 44.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*, 1–17.

Makowski, D., Ben-Shachar, M. S., Chen, S. H., & Lüdecke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in Psychology*, *10*, 498833.

Maxwell, N. P., & Huff, M. J. (2022a). Is discriminability a requirement for reactivity? Comparing the effects of mixed vs. pure list presentations on judgment of learning reactivity. *Memory & Cognition*, 1–16.

Maxwell, N. P., & Huff, M. J. (2022b). Reactivity from judgments of learning is not only due to memory forecasting: Evidence from associative memory and frequency judgments. *Metacognition and Learning*, *17*, 589–625.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*, 159–163.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*, 174–179.

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, *145*, 200–219.

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, *20*, 378–384.

Murphy, D. H., & Castel, A. D. (2021). Metamemory that matters: Judgments of importance can engage responsible remembering. *Memory (Hove, England)*, *29*, 271–283.

Murphy, D. H., Hoover, K. M., & Castel, A. D. (2023). Strategic metacognition: Self-paced study time and responsible remembering. *Memory & Cognition*, *51*, 234–251.

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition*, *48*, 745–758.

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*, 1170–1187.

Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Kreuger, S. (2005). Divided attention in younger and older adults: Effects of Strategy and relatedness on memory performance and secondary Task costs. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*, 520–537.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*, 102–116.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the labor-in-vain effect. *Journal of Experimental Psychology: Learning Memory and Cognition*, *14*, 676–686.
Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, *26*, 125–173.
Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, *5*, 207–213.
Park, D. C. (2000). The basic mechanisms accounting for age-related decline in cognitive function. *Cognitive Aging: A Primer*, *11*, 3–19.
Park, D. C., & Festini, S. B. (2017). Theories of memory and aging: A look at the past and a glimpse of the future. *Journals of Gerontology: Psychological Sciences*, *72*, 82–90.
Rhodes, M. G. (2016). Judgments of learning. In J. Dunlosky, & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). Oxford University Press.
Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology*, *46*, 168–175.
Rivers, M. L., Janes, J. L., & Dunlosky, J. (2021). Investigating memory reactivity with a within-participant manipulation of judgments of learning: Support for the cue-strengthening hypothesis. *Memory (Hove, England)*, *29*, 1342–1353.
Roberts, J. R., & Maxfield, M. (2021). *A 2-study psychometric evaluation of the modified dementia worry scale* (Vol. 36, p. 1533317521995322). American Journal of Alzheimer's Disease & Other Dementias®.
Robey, A. M., Dougherty, M. R., & Buttaccio, D. R. (2017). Making retrospective confidence judgments improves learners' ability to decide what not to study. *Psychological Science*, *28*, 1683–1693.
Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*, 140–144.
Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, *16*, 754–760.
Salthouse, T. A. (2019). Trajectories of normal cognitive aging. *Psychology and Aging*, *34*, 17–24.
Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory & Cognition*, *49*, 955–967.
Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*, 553–558.
Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*, 204–221.
Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, *5*, 315–316.
Spreng, R. N., & Turner, G. R. (2019). The shifting architecture of cognition and brain function in older adulthood. *Perspectives on Psychological Science*, *14*, 523–542.
Staudinger, U. M., Cornelius, S. W., & Baltes, P. B. (1989). The aging of intelligence: Potential and limits. *The Annals of the American Academy of Political and Social Science*, *503*, 43–59.
Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgments of retention (JORs). *The Quarterly Journal of Experimental Psychology*, *65*, 1376–1396.
Tauber, S. K., & Witherby, A. E. (2019). Do judgments of learning modify older adults' actual learning? *Psychology and Aging*, *34*, 836–847.
Tendeiro, J. N., Kiers, H. A., Hoekstra, R., Wong, T. K., & Morey, R. D. (2024). Diagnosing the misuse of the Bayes factor in Applied Research. *Advances in Methods and Practices in Psychological Science*, *7*, 25152459231213371.
Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*, 1024–1037.
Thiede, K. W., Anderson, C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*, 66–75.
Thomas, A. K., & Gutchess, A. (Eds.). (2020). *The Cambridge Handbook of Cognitive Aging: A Life Course Perspective*. Cambridge University Press.
Umanath, S., & Marsh, E. J. (2014). Understanding how prior knowledge influences memory in older adults. *Perspectives on Psychological Science*, *9*, 408–426.
Undorf, M., Schäfer, F., & Halamish, V. (2024). *Making judgments of learning either enhances or impairs memory: Evidence from 17 experiments with related and unrelated word pairs* (Vol. 10, p. 117108). Psychology.
Witherby, A. E., & Tauber, S. K. (2017). The influence of judgments of learning on long-term learning and short-term performance. *Journal of Applied Research in Memory and Cognition*, *6*, 496–503.