American Psychological Association

Psychology and Aging

0882-7974/18/$12.00

2018, Vol. 33, No. 3, 497–511
http://dx.doi.org/10.1037/pag0000256

Unitization of Word Pairs in Young and Older Adults: Encoding Mechanisms and Retrieval Outcomes

Siri-Maria Kamp
Trier University

Regine Bader and Axel Mecklinger
Saarland University

We investigated whether healthy older adults are able to use an episodic encoding strategy known as *unitization*, which allows for subsequent associative retrieval based on familiarity, to overcome their associative memory deficit. Young and healthy older participants were presented with word pairs either together with a definition that allowed to combine the word pairs to a new concept (high unitization condition), or together with a sentence frame (low unitization condition). In Experiment 1, an age-related reduction in performance on a standard associative recognition test was observed in both conditions. This deficit was unexpectedly not reduced, but tended to be larger in the high than the low unitization condition. According to receiver-operating characteristics, this difference was due to a reduction of recollection, but not familiarity, in the high unitization condition. Instead of a standard recognition test, Experiment 2 used a 2 alternative forced choice (2AFC) test designed to maximize the contribution of familiarity to associative recognition. Although the disadvantage of older adults in the high versus the low unitization condition was abolished, there was still no performance advantage for the high unitization condition. Event-related potentials (ERPs) recorded during the encoding phase of Experiment 1 suggest that, while young adults engage in predictive processing during unitization, older adults do not engage in such predictive processing, which may prevent them from using unitization to their advantage in the subsequent associative memory test. We discuss the task characteristics that have an impact on the effect of unitization conditions on associative memory in older adults.

Keywords: associative memory, unitization, episodic encoding, healthy aging, event-related potentials

Supplemental materials: http://dx.doi.org/10.1037/pag0000256.supp

As humans age, their episodic memory declines, with a particularly strong decline of memory for associations between stimuli, compared with item memory for individual stimuli (Old & Naveh-Benjamin, 2008). A special strategy to encode associations, known as *unitization,* is to mentally integrate the components of an association into a single, unified memory representation (Graf & Schacter, 1989). Essentially, this creates an item representation from associative information and may therefore allow older adults to use their relatively intact item memory to compensate for their reduced associative memory. Prior studies testing this idea have provided mixed results. In one study (Bridger et al., 2017), older adults were not able to utilize unitization encoding to their advantage to the same extent as young adults when unitization was based on a spatial arrangement of two objects that made an interaction between them plausible (vs. implausible). Here, we explored whether older adults would be successful in applying a different unitization strategy that establishes a semantically meaningful relationship between the components of an association, perhaps helping them overcome the associative memory deficit. Furthermore, we explored whether aging affects the unitization process itself at the time of encoding by examining event-related potentials (ERPs) during encoding.

Unitization as a Strategy to Encode Associations in Aging

In addition to a reduced ability to bind together new, arbitrary associations, older adults may exhibit a deficit in applying strategies that aid the effective formation of relational memory traces (Shing, Werkle-Bergner, Li, & Lindenberger, 2008). Indeed, older adults are less likely than young adults to spontaneously form sentences with, or mental images of, the components of a to-be-encoded association (Naveh-Benjamin, Brav, & Levy, 2007). If strategic difficulties contribute to the associative deficit, providing

Siri-Maria Kamp, Department of Psychology, Trier University; Regine Bader and Axel Mecklinger, Department of Psychology, Saarland University.

This research was supported by the International Research Group "Adaptive Minds" (IRTG 1457) funded by the German Research Foundation. During this work, Siri-Maria Kamp was supported by a postdoctoral position in the IRTG. Parts of the data from the young adults of the present article were reported in previous journal articles (Kamp et al., 2016, 2017). Preliminary results from both age groups were presented at the Aging and Cognition Conference in Dortmund, Germany, in April 2015. We thank Ricarda Endemann, Hong H. Pham, and Jana Döring for their assistance with data collection, and Emma K. Bridger for discussions of previous versions of this work.

Correspondence concerning this article should be addressed to Siri-Maria Kamp, Department of Psychology, Trier University, 54290 Trier, Germany. E-mail: kamp@uni-trier.de

support in the form of strategy instructions or training may help older adults compensate for their binding deficit. Evidence supporting this idea comes from a study in which young and older participants were instructed to generate a sentence to encode word pairs, while in the control condition no strategy instructions were given (Naveh-Benjamin et al., 2007). However, in another study (Healy, Light, & Chung, 2005), associative memory of young, but not older adults benefitted from similar instructions. Furthermore, another study reported that although associative memory of all age groups benefitted from training in an interactive imagery encoding strategy, age differences were *enhanced* after training due to a larger improvement in young adults (Shing et al., 2008). Perhaps the inconsistency of these results is due to a relatively high amount of self-initiation inherent in some strategies, such as when participants have to generate a sentence on their own, because older adults usually perform much better when more environmental or strategic support is available (e.g., Craik & Rose, 2012). Indeed, Luo, Hendriks, and Craik (2007) demonstrated that age differences in memory performance are reduced when encoding conditions foster environmental support, but fail to do so or even increase the age difference when encoding conditions designed to foster deeper encoding rely on self-initiated control processes. Unitization as an encoding strategy may impose lower demands on self-initiated top-down guided strategic processing during encoding and/or retrieval and provide environmental or schematic support, and may therefore be of higher utility to reduce the age-related associative memory deficit (see Bridger et al., 2017, for a similar argument).

Unitization reduces the extent to which the different components of an association remain individual, separate entities flexibly bound together by a relational trace. Unlike the strategies reviewed above, unitization leads to a greater relative reliance of retrieval on familiarity than recollection (e.g., Bader, Mecklinger, Hoppstädter, & Meyer, 2010; Bridger et al., 2017; Diana, Yonelinas, & Ranganath, 2008; Jäger, Mecklinger, & Kipp, 2006; Rhodes & Donaldson, 2008; Tibon, Gronau, Scheuplein, Mecklinger, & Levy, 2014; Yonelinas, Kroll, Dobbins, & Soltani, 1999). Familiarity is an automatic process that is independent of hippocampal activity (Yonelinas, 2002) and is relatively unaffected by aging (Healy et al., 2005; Light, Prull, La Voie, & Healy, 2000). By contrast, recollection is an effortful, hippocampus-dependent retrieval mode (Yonelinas, 2002) with which older adults have difficulties. Because retrieval of associations usually requires recollection, but unitization encoding allows for retrieval of associations via familiarity, older adults should have smaller difficulties recovering associations when they are unitized.

To explore this idea, Bastin et al. (2013) instructed participants to unitize objects with background colors by imagining each object in the respective color. In a nonunitization control condition, older adults showed reduced memory for the object-color pairings relative to young adults, but in the unitization condition, this associative deficit was absent when participants indicated that the color-object pairing was readily unitizable. Comparing preexisting (i.e., preexperimentally unitized) compound words with word pairs that are grammatically legal, yet nonexisting (i.e., nonunitized) compound words, two other studies (Ahmad, Fernandes, & Hockley, 2015; Zheng, Li, Xiao, Broster, & Jiang, 2015) showed a reduction of the age-related associative memory deficit for the preexisting compound words. Following a similar idea, Delhaye and Bastin (2018) also asked participants to encode either preexisting compound words or noncompound words. However, although with a two alternative forced choice (2AFC) recognition test the associative deficit of older adults was abolished, there was no disproportionate benefit for older adults in the compound word condition. The results of this study are therefore somewhat in conflict to the former two studies. Nevertheless, taken together, there are several studies that underline the potential of unitization to reduce the associative memory deficit in old age. The unitization tasks reviewed so far, however, are not without limitations. For example, all objects in Bastin et al. (2013) were paired with one of two colors, whereas in a more natural situation, associations consist of more arbitrary pairings. On the other hand, contrasting compound words to noncompound words does not tap into strategic aspects of unitization encoding, because unitization is inherent in the study material in the form of high preexperimental associative strength of the components.

Two other studies reported markedly different result patterns. Jäger, Mecklinger, and Kliegel (2010) asked participants to encode pairs of pictures of faces, which either belonged to the same or two different individuals. When the pictures depict the same individual, unitization of the pair is facilitated in young adults (Jäger et al., 2006). However, older adults showed disproportionately *reduced* associative memory performance in this unitization condition. The authors speculated that due to age-related changes in the medial temporal lobes, unitization encoding may be difficult for older adults when the feature overlap between the to-be-associated stimuli is high. Bridger et al. (2017) used a paradigm with low feature overlap of stimulus pairs. Unitization was either facilitated or impeded based on a spatial arrangement that made an interaction between two semantically unrelated objects plausible or implausible. Memory of both young and older adults benefitted from the plausible arrangement, but the effect was more pronounced for young adults. By this, contrary to the authors' predictions unitization increased the age differences in associative memory performance. Of note, an early frontal ERP effect generally associated with familiarity was present in both age groups, suggesting that familiarity-based associative memory is intact in old adults. However, ERP effects did not vary with spatial plausibility. In this study, unitization of stimulus pairs did not create a semantically meaningful relationship between the components of an association. According to the authors, associative memory of older adults may only benefit from unitization manipulations when their increasing reliance on semantic relations (see Ofen & Shing, 2013) is taken into account.

To test whether a unitization task that creates new semantic relationships between the to-be-unitized stimulus pairs improves associative memory in older adults, in the present study we utilized a paradigm in which a novel conceptual unit is formed with the different parts of an association (Quamme, Yonelinas, & Norman, 2007). Preexperimentally unrelated word pairs (e.g., "family" and "bikini;" see Figure 1) are encoded in one of two conditions that promote unitization of the pair into a holistic concept to different extents. In the "high unitization" condition, each word pair is presented together with a definition (e.g., "A piece of bathing attire that is shared between cousins") that provides an entity-defining framework to tie together the pair into a coherent new concept. The pair therefore becomes a new, meaningful compound word. In the "low unitization" condition, the pair is presented together with a sentence frame in which the two words can be inserted (e.g., "In

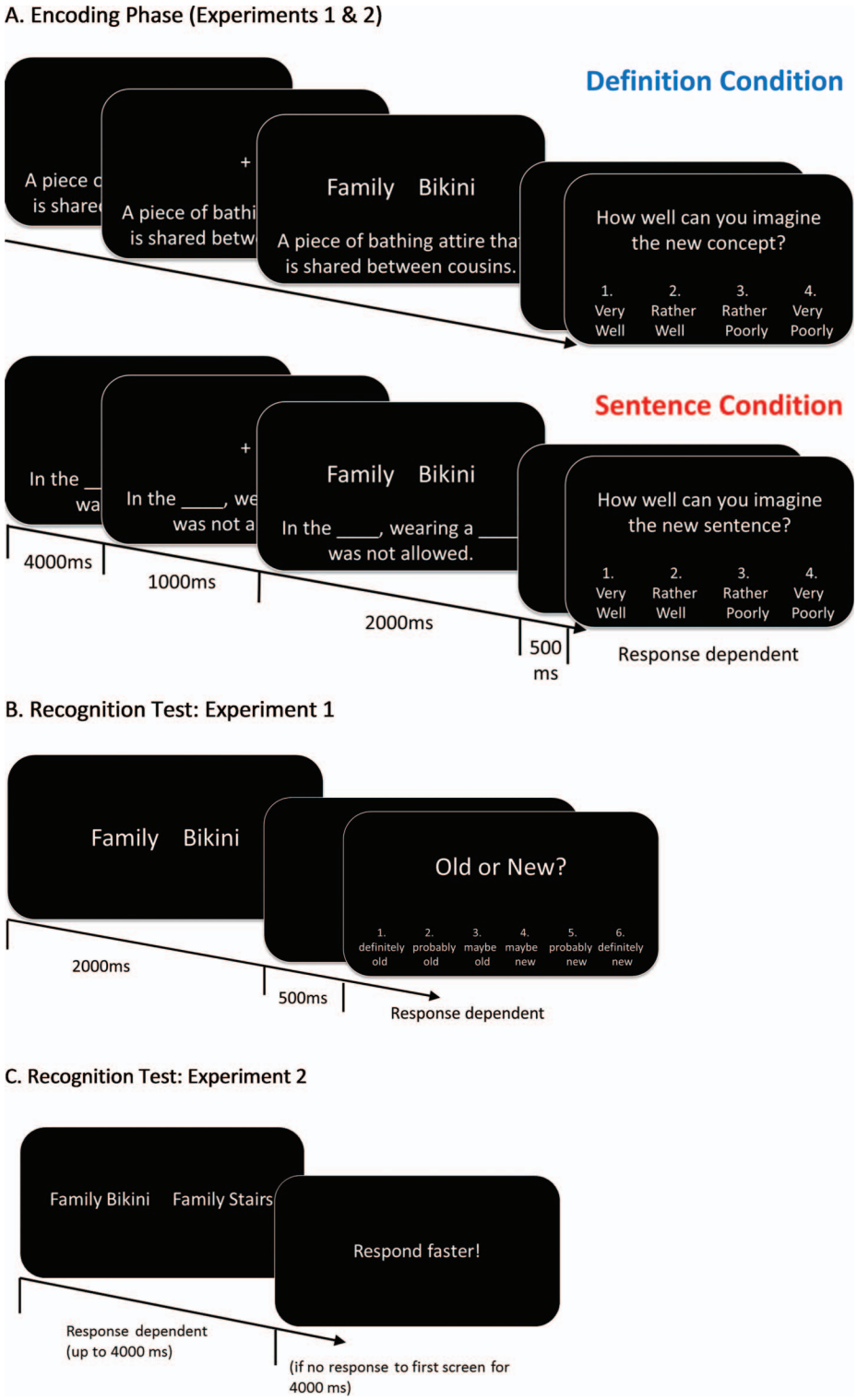

Figure 1. Experimental design: Trial structure at encoding and recognition. (A) Participants were randomly assigned to either the definition or the sentence encoding condition. The encoding phase was identical in Experiments 1 and 2. (B and C) The test structure (yes/no recognition in Experiment 1, 2AFC recognition in Experiment 2) was different in Experiments 1 (B) and 2 (C), but was identical for participants in both conditions within each experiment. In both the encoding and test phase, and in both experiments, the intertrial interval (from the offset of the rating screen to the onset of the next sentence/definition or the next word pair, respectively) was 2,000 ms. Example stimuli have been translated from German to English. See the online article for the color version of this figure.

the ___, wearing a ___ was not allowed"). In young adults, retrieval of word pairs that are encoded in the high unitization condition relies relatively more strongly on familiarity than on recollection, compared with the control condition (Bader et al., 2010; Kamp, Bader, & Mecklinger, 2016).

This paradigm has been well-studied in young adults and examining its benefit to associative memory in older adults promises important new insights into the subject matter. First, the word pairs are preexperimentally unrelated and unassociated, so an active unitization process is required during encoding. Second, unitization relies on the creation of a semantic relationship between stimuli. Because memory in the aging brain becomes increasingly reliant on preexisting semantic knowledge (Ofen & Shing, 2013), this strategy may be beneficial to older adults. Finally, the availability of the definition to generate the novel concept from the word pair should provide ample schematic support, contributing to a successful unitization process in older adults.

Age Differences in ERPs During Episodic Encoding

Bridger et al. (2017) examined ERP correlates of recollection and familiarity during the recognition test, but no prior studies have explored the neurocognitive mechanisms underlying age differences in the formation of unitized versus nonunitizes representations during the encoding phase. To this end, a useful approach is to examine the extent to which ERP amplitudes covary with subsequent associative retrieval success ("subsequent memory effects" [SMEs]; Karis, Fabiani, & Donchin, 1984; Paller & Wagner, 2002). Comparing ERP SMEs between young and older adults, Friedman, Ritter, and Snodgrass (1996) and Kamp and Zimmer (2015) found reliable ERP SMEs in young, but not older adults, and concluded that older adults utilized elaborative encoding strategies to a lesser extent. Cansino, Trejo-Morales, and Hernández-Ramos (2010) interpreted age differences in onset and distribution of ERP SMEs to index differences in the efficiency of source memory encoding. Two other studies reported that SMEs in young, but not older adults varied depending on the subsequent categorization into remember versus know responses (Friedman & Trott, 2000) or subsequent confidence judgments (Gutchess, Ieuji, & Federmeier, 2007), suggesting that young, but not older participants engaged differential encoding mechanisms. Together, these results demonstrate the utility of ERP SMEs to examine age differences in encoding mechanisms.

A previous analysis (Kamp, Bader, & Mecklinger, 2017) of ERP SMEs for the young adults in the present study revealed that an early SME with contributions from parietal and frontal areas was unique to the definition condition, suggesting that it plays a role in the formation of a new, holistic concept. A later frontal SME was found in both conditions, suggesting that it plays a more general role in associative encoding. Young adults therefore employ at least partially different encoding operations in the unitization versus control condition (see also Haskins, Yonelinas, Quamme, & Ranganath, 2008). In the present study, we compared ERPs during encoding between young and older adults.

Experiment 1

We investigated whether unitization of word pairs through use of a definition that ties the pair into a new conceptual unit affects the magnitude of the associative memory disadvantage of older adults. We predicted a performance advantage of older adults in the definition over the sentence condition and a decreased age difference in associative memory performance through unitization. We also examined receiver-operating characteristics (ROC) of recognition ratings given during the test phase (e.g., Yonelinas, 1994), which exhibit distinct shapes depending on the contribution of recollection and familiarity to recognition memory. Accordingly, parameters can be estimated to quantify the two processes. We expected that retrieval of both young and older adults relies more strongly on familiarity in the definition than in the sentence condition. Finally, we explored whether older (like young) adults engage different encoding processes in the two conditions by analyzing ERP SMEs during the encoding phase. Note that parts of the data from the young adults are presented in detail elsewhere (Kamp et al., 2017), but will be included in the present article as a control group for the sample of older adults.

Method

The procedures of all experiments were approved by a local ethics committee. All participants provided informed consent.

Participants. Forty-two community-dwelling older (65- to 79-years-old) and 42 younger (up to 30-years-old) native German speakers were paid €8 for their participation (see Table 1). All participants reported to be healthy and free of neurological conditions, had normal or corrected-to-normal vision, and were randomly assigned to either the high or the low unitization encoding condition. At the end of the experimental session or in a separate session within 3 months of the experiment we administered the digit-symbol substitution test indexing processing speed (Wechsler, 2008), the spot-a-word test to measure vocabulary (Lindenberger, Mayr, & Kliegl, 1993) and a computerized counting span task to measure working memory (Conway et al., 2005; Table 1). To explore whether our sample shows the typical age-related differences in these neuropsychological test measures and to ensure that the participants assigned to the two encoding conditions did not differ in these measures, we conducted separate 2 (age group) $\times$ 2 (encoding condition) ANOVAs on each measure. All three ANOVAs revealed significant age differences (all p *value*s $< .05$), but no main effects or interactions of condition (all p *value*s $> .26$). Young adults outperformed older adults in all tests except for the spot-a-word test, in which the older participants demonstrated better performance. This is in line with an abundance of prior evidence suggesting that some aspects of cognitive function, including working memory and processing speed, decline during aging, while others, such as semantic memory and vocabulary, are relatively unaffected. All older adults also completed the mini mental state examination (MMSE; Folstein, Folstein, & McHugh, 1975) and scored at least 27 points, so there was no evidence for cognitive impairments in our sample.

Participants were excluded from parts of the behavioral analysis if they did not use the entire rating scale during encoding and from the ERP analysis if they provided less than 10 artifact-free trials in one of the ERP averages. The Results section contains further detail on data exclusion.

Stimuli and procedure. The session began with the preparations for the EEG recording, which took up to 45 min. The stimuli and procedure were a modification of Bader, Mecklinger,

Table 1
Demographic Data and Neuropsychological Test Measures (Mean ± SD) for Each Participant Group

Measure	Experiment 1				Experiment 2	
	Young		Old		Old	
	Definition (n = 21)	Sentence (n = 21)	Definition (n = 21)	Sentence (n = 21)	Definition (n = 22)	Sentence (n = 22)[4]
Demographics						
Age**	23.1 (2.4)	23.5 (2.9)	73 (3.7)	71.5 (3.1)	73.4 (4.1)	74.8 (3.8)
Gender (# females)	12	14	10	8	11	11
Neuropsychological tests						
Digit symbol[1]**	44.9 (7.5)	47.3 (6.7)	31.7 (4.1)	29.2 (4.9)	29.1 (7.8)	27.2 (6.8)
MSWT[2]**	22.6 (5.0)	22.7 (2.9)	28.3 (3.3)	28.0 (3.1)	27.3 (3.9)	27.9 (2.7)
Counting span[3]**	38.1 (6.7)	38.3 (7.0)	33 (8.1)	31.8 (7.5)	32.3 (7.6)	28.8 (8.0)

[1] Number of positions filled in 60 s of time, maximum value is 67. [2] Number of correct identifications of the word, out of 35. [3] Number of positions correctly filled, out of 54 positions total. [4] Data from the neuropsychological tests were lost for one participant, so the means and standard deviations are based on n = 21.
** $p < .01$: Significant age difference (t-test collapsed across conditions and, for old adults, across experiments. See main text for further statistical analyses).

Hoppstädter, and Meyer (2010; see also Bader, Opitz, Reith, & Mecklinger, 2014), which allowed for the examination of ERP SMEs (Kamp et al., 2017). Participants were presented with word pairs, either together with a definition that provided a basis to integrate the words into a compound word (high unitization or definition condition), or with a sentence including two blank spaces in which the words could be mentally inserted (low unitization or sentence condition). All words included in the pairs were nouns of medium frequency (10–500 occurrences per million) and were suitable for German compound construction. All pairs were judged as neither preexperimentally semantically related nor thematically associated in a pilot rating study conducted by Bader et al. (2010, 2014). The task during the encoding phase was to evaluate the plausibility of each concept or sentence on a scale of 1 (*very well*) to 4 (*very poorly*; the scale was reversed in a counterbalanced manner). Although age differences in associative memory are typically largest under intentional encoding instructions (Old & Naveh-Benjamin, 2008), we here used an incidental encoding task: No mention of the recognition test was made. The rationale for this was to prevent both age groups from using self-initiated encoding strategies other than the formation of the new concept or sentence. Furthermore, the prior studies that used the same experimental paradigm also used incidental encoding, so use of an incidental encoding task made our experiment comparable with previously published studies in young adults.

An encoding trial (Figure 1A) began with the presentation of the sentence/definition alone for 4 s, followed by the additional presentation of a fixation cross in the center of the screen for 1 s and then the word pair for 2 s. The sentence/definition remained on the screen together with the word pair. Next, after a 500-ms blank screen, the rating screen was shown. The response was provided with a serial response box and terminated the rating screen. The intertrial interval consisted of a fixation cross, shown for 2 s. In total, 160 experimental trials were encountered in the encoding phase. After a distracter task (counting backward in steps of three for 5 min), participants were presented in pseudorandom order with 80 word pairs that were exactly the same as in the encoding phase and were therefore to be endorsed as "old," and 80 word pairs that were recombinations of two words from different pairs, which were to be judged as "new" (referred to as "recombined" in this article). This judgment was provided on a 6-point scale from *definitely old* to *definitely new* (Figure 1B). Note that this test format relies on associative memory for the particular pairing of words, but not on memory for the individual words, because in old and recombined pairs alike, both words are previously studied. A recognition trial consisted of the presentation of the word pair for 2 s. After a blank screen for 500 ms, the rating scale was shown, which was terminated by the participants' response. Between the offset of the rating screen and the onset of the next word pair a fixation cross was shown for 2,000 ms. Both the encoding and the recognition phase were preceded by several practice trials, which could be repeated if needed. Throughout the experiment, self-paced breaks were allowed after each set of 40 trials.

Behavioral data analysis. We collapsed across all "old" judgments (*definitely old*, *probably old*, and *maybe old*) to calculate Pr scores (proportion of hits minus proportion of false alarms). As unitization should be affected by the quality of integration into the new concept, we separated trials that at encoding had received a *very well* or *rather well* judgment (high fit trials) from those that had received a *rather poorly* or *very poorly* judgment (low fit trials; see also Bastin et al., 2013) and included "fit rating" as a factor in our statistical analyses (see also Kamp et al., 2016). Note that this separation into high fit and low fit trials was not possible for the ROC and SME analyses (see below) because a higher number of trials is needed for these analyses and a subdivision would not have left enough trials in each stimulus class.

Estimates of familiarity and recollection were calculated for each participant by fitting a curve to the empirical ROC using the excel solver function (Dodson, Prinzmetal, & Shimamura, 1998), analogously to Jäger and Mecklinger (2009). Thus, we estimated familiarity (d'), recollection of intact pairs (Ro) and recollection of recombined pairs (Rn). The recollection parameters were restricted to values between 0 and 1, and the familiarity parameter to values above 0 (see also Parks & Yonelinas, 2015).

EEG recording and analysis. The EEG was recorded from 28 Ag/AgCl scalp electrodes embedded according to the 10–20 electrode system in an elastic cap with FCz as the ground electrode. An additional four electrodes were used to record horizontal

and vertical eye movements and blinks. The EEG was amplified from DC to 250 Hz using BrainAmp (Brain Products, Inc.) DC amplifiers and Brain Vision Recorder software and was digitized at 500 Hz. The online reference was the left mastoid electrode and we rereferenced the EEG to linked mastoids off-line.

We filtered the EEG at 0.1–20 Hz and extracted segments from 200 ms before to 2,000 ms after word pair onset in the encoding phase. We used the modified regression-based analysis (Gratton, Coles, & Donchin, 1983) implemented in Brain Vision Analyzer 2.0 to correct for EOG activity. After rejection of residual artifacts we calculated subject ERPs for trials associated with subsequent high confidence hits (referred to here as "subsequent hits") and trials associated with subsequent medium- or low-confidence hits, or misses ("subsequent misses"). The baseline correction used the 200 ms that preceded word pair onset. Mean amplitudes were analyzed for nine electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4) and for an early (300 ms–600 ms after word pair onset) and a late (1,200 ms–2,000 ms) time window (Kamp et al., 2017).

Statistical analysis. Using IBM SPSS software, we analyzed the data in mixed ANOVAs as well as paired and independent samples *t* tests. Greenhouse-Geisser corrected degrees of freedom and *p* values are reported where appropriate. Follow-up tests comprised lower level ANOVAs and *t* tests.

Behavioral Results

Table 2 and Figure 2 show the behavioral results from Experiment 1.

Recognition performance. The inclusion of the "fit rating" factor in the analysis of Pr scores necessitated the exclusion of the data from one older participant in each condition, because they did not judge any encoding trials as poorly imaginable.[1] This resulted in 20 participants in each condition for the older and 21 participants in each condition for the young adults. A mixed ANOVA with the between-subject factors age (young vs. old) and condition (definition vs. sentence) and the within-subject factor "fit rating" (high vs. low fit) on Pr scores revealed a main effect for age, $F(1, 78) = 19.28$, $p < .001$, $\eta_p^2 = .20$, and a main effect for condition, $F(1, 78) = 5.77$, $p = .02$, $\eta_p^2 = .07$. The age by condition interaction only approached significance, $F(1, 78) = 3.09$, $p = .08$, $\eta_p^2 = .04$. Furthermore, there was a main effect for fit rating, $F(1, 78) = 21.53$, $p < .001$, $\eta_p^2 = .22$, qualified by an interaction between fit rating and condition, $F(1, 78) = 8.60$, $p = .004$, $\eta_p^2 = .10$. Importantly, although we had predicted an age by condition interaction, numerically the interaction was opposite to our prediction: Older adults tended to show *lower* recognition performance in the definition than in the sentence condition. To explore this unexpected pattern further, we followed-up the nonsignificant interaction by separate analyses for each age group. It should be noted that replications of the nonsignificant interaction with higher power are necessary, and the results of these follow-up analyses of a nonsignificant interaction should be regarded with caution.

For the young adults, the condition by fit rating ANOVA revealed no difference in Pr scores between the two conditions ($p > .65$; Kamp et al., 2017). Analogously to the overall ANOVA, there was a significant main effect for fit rating, $F(1, 40) = 19.60$, $p < .001$, $\eta_p^2 = .33$, qualified by a fit rating by condition interaction, $F(1, 40) = 4.25$, $p = .05$, $\eta_p^2 = .10$. In the definition condition, recognition performance was better for word pairs for which participants had indicated that they could imagine the new concept well, $t(20) = 5.19$, $p < .001$, $d = 1.13$, but in the sentence condition, fit rating did not significantly affect performance, $t(20) = 1.52$, *ns*.

For the older adults, a significant main effect for condition revealed that Pr scores were lower in the definition, compared with the sentence condition, $F(1, 38) = 9.19$, $p = .004$, $\eta_p^2 = .20$. Furthermore, there was a main effect for fit rating, $F(1, 38) = 5.51$, $p = .024$, $\eta_p^2 = .13$, qualified by a fit rating by condition interaction, $F(1, 38) = 4.35$, $p = .044$, $\eta_p^2 = .10$. Like for the young adults, only in the definition condition were high fit trials associated with better recognition performance than low fit trials, $t(19) = 3.03$, $p < .001$, $d = .68$ (sentence condition: $p > .85$). Notably, performance in the definition condition still tended to be reduced relative to the sentence condition when only the high fit trials were considered ($p = .052$), so it is unlikely that the lower performance in this condition was driven by trials in which participants could not readily imagine the new concept.

The results were analogous when hit rates rather than Pr scores were analyzed (see Table 2). However, a 2 (age) × 2 (condition) ANOVA on the false alarm rates revealed a main effect for age, $F(1, 78) = 4.35$, $p = .002$, $\eta_p^2 = .12$, but no main effect for condition or age by condition interaction (both *p values* > .33).

ROC analysis. The 2 (age) × 2 (condition) ANOVA on the familiarity parameter (Figures 2B and 2C) revealed a main effect for age, $F(1, 80) = 35.06$, $p < .001$, $\eta_p^2 = .31$, but no other main or interaction effect (both *p values* > .49). Older adults generally exhibited lower familiarity parameters than young adults.

For the recollection parameter (Ro), the age by condition ANOVA revealed two main effects—age: $F(1, 80) = 7.32$, $p = .008$, $\eta_p^2 = .84$; condition: $F(1, 80) = 4.57$, $p = .036$, $\eta_p^2 = .05$. Older adults exhibited a lower Ro parameter than did young adults. Furthermore, Ro was smaller in the definition, compared to the sentence, condition. Although inspection of Figure 2C suggests that the difference between conditions in the Ro parameter was larger for the older, compared with the young adults, the interaction of age and condition was not significant ($p > .16$).

Estimates of the second recollection parameter (Rn) were zero for 52 of the 84 participants, and were generally small (Figure 2C, Table 2). The age by condition ANOVA did not reveal any main effects or interactions (all *p* values > .68).

Behavioral results: Summary. Older adults showed a general reduction in associative memory, compared with young adults. However, in contrast to our hypothesis, this age difference was not reduced, but tended to be *increased* in the definition, compared with the sentence condition. ROC analyses revealed reductions in both the familiarity and the recollection parameter in the older adults. There were no differences between the two conditions in familiarity, but the recollection parameter (Ro) was reduced in the definition compared with the sentence condition across both age groups.

[1] Note that an analysis including these participants, which did not include the factor fit rating, led to analogous result patterns for the main effects and interactions of age and condition.

Table 2
Means (±SD) of the Behavioral Data from the Test Phases of Experiments 1 and 2

	Experiment 1				Experiment 2	
	Young		Old		Old	
Measure	Definition (n = 21)	Sentence (n = 21)	Definition (n = 21)	Sentence (n = 21)	Definition (n = 22)	Sentence (n = 22)
All trials						
False alarm rate	.15 (.11)	.15 (.14)	.29 (.17)	.24 (.15)		
Hit rate	.76 (.14)	.79 (.10)	.63 (.17)	.72 (.09)		
Pr score	.61 (.23)	.64 (.19)	.34 (.23)	.48 (.18)		
Percent correct					.65 (.11)	.65 (.10)
Familiarity (d')	1.04 (.65)	1.12 (.66)	.30 (.43)	.39 (.52)		
Recollection (Ro)	.51 (.20)	.54 (.18)	.32 (.25)	.48 (.20)		
Recollection (Rn)	.10 (.20)	.08 (.22)	.07 (.11)	.08 (.16)		
"High fit" trials						
Hit rate	.81 (.13)	.80 (.10)	.66 (.17)[1]	.73 (.10)[1]		
Pr score	.66 (.22)	.65 (.19)	.38 (.23)[1]	.51 (.18)[1]		
Percent correct					.67 (.12)	.64 (.09)[2]
"Low fit" trials						
Hit rate	.67 (.17)	.75 (.16)	.55 (.19)[1]	.72 (.15)[1]		
Pr score	.52 (.26)	.60 (.21)	.26 (.23)[1]	.50 (.20)[1]		
Percent correct					.63 (.16)	.66 (.12)[2]

[1] Data are based on 20 participants only because one participant provided no "low fit" trials. [2] Data are based on 21 participants only because one participant provided no "low fit" trials.

ERP Results

Data were excluded from the ERP analysis when less than 10 artifact-free EEG trials were available either for the "subsequent hit" or "subsequent miss" trial type, resulting in 17 and 18 young, and 17 and 16 older, participants in the definition and the sentence condition, respectively. It is worth noting that the low recognition performance of several older participants in the definition condition led to a very small trial number in the "subsequent hit" category in these participants, resulting in their exclusion from the ERP analysis. By contrast, for the older adults in the sentence condition and the young adults in both conditions, exclusion from the ERP analysis was not selective to low performers—the participants excluded from these groups rather had a high amount of artifacts from movements and other sources. Exclusion of the low performers only in the old/definition group resulted in an ERP sample that did not differ significantly in their behavioral performance from older adults in the sentence condition. It is therefore important to keep in mind that the result patterns in ERPs cannot be attributed to performance differences within the older age group.

The grand average ERPs from a frontal (Fz) and a parietal (Pz) electrode are shown in Figure 3a, and the distributions of the subsequent memory effects (subsequent hits-subsequent misses) are shown in Figure 3b. We analyzed mean amplitudes in 3 (Anteriority: frontal, central, parietal) × 3 (Laterality: left, central, right) × 2 (Subsequent Memory: subsequent hit, subsequent miss) × 2 (Condition: sentence, definition) × 2 (Age Group: young, old) mixed ANOVAs and report only main effects and interactions involving the factors subsequent memory, age, or condition.

Early time window (300 ms–600 ms). In the overall ANOVA, anteriority interacted with condition, $F(1.17,75.07) = 7.19$, $p = .006$, $\eta_p^2 = .10$, and age, $F(1.17,75.07) = 7.12$, $p = .007$, $\eta_p^2 = .10$. There was also a three-way interaction between laterality, condition, and age, $F(1.75,111.87) = 3.41$, $p = .04$, $\eta_p^2 = .05$. There was a significant main effect for subsequent memory, $F(1, 64) = 10.87$, $p = .002$, $\eta_p^2 = .15$. The three-way interaction between subsequent memory, condition, and age only approached significance, $F(1, 64) = 3.65$, $p = .06$, $\eta_p^2 = .05$. Finally, there was a main effect for condition, $F(1, 64) = 6.10$, $p = .02$, $\eta_p^2 = .09$, qualified by a condition by age interaction, $F(1, 64) = 11.74$, $p = .001$, $\eta_p^2 = .16$. Due to this significant interaction, we performed separate lower-level 3 (Anteriority: frontal, central, parietal) × 3 (Laterality: left, central, right) × 2 (Subsequent Memory: subsequent hit, subsequent miss) × 2 (Condition: sentence, definition) mixed ANOVAs for each age group separately.

As we have reported elsewhere, for the young adults an early SME was prominent exclusively in the definition, but not in the sentence condition (Kamp et al., 2017; see also Figures 3a and 3b). The ANOVA for the older adults revealed a significant SME, $F(1, 31) = 6.82$, $p = .01$, $\eta_p^2 = .18$, but there were no interactions involving the factors condition and subsequent memory (all p *values* $> .55$). The conclusion that for the older adults the two conditions did not differ in the early SME was further confirmed in pairwise contrasts of SME magnitude at each electrode (all p *values* $> .18$). There was also an interaction between anteriority and condition, $F(1.18,36.45) = 4.33$, $p = .04$, $\eta_p^2 = .12$, reflecting the more pronounced frontal negativity elicited for the older adults in the definition condition.

Late time window (1,200 ms–2,000 ms). The overall ANOVA revealed a main effect for subsequent memory, $F(1, 64) = 14.74$, $p < .001$, $\eta_p^2 = .19$, but no interaction involving the factors condition or age (all p values $> .24$). As shown previously, for the young adults this slow wave SME was evident in both conditions (Kamp et al., 2017). The lack of significant interactions suggests that there were no condition differences between the

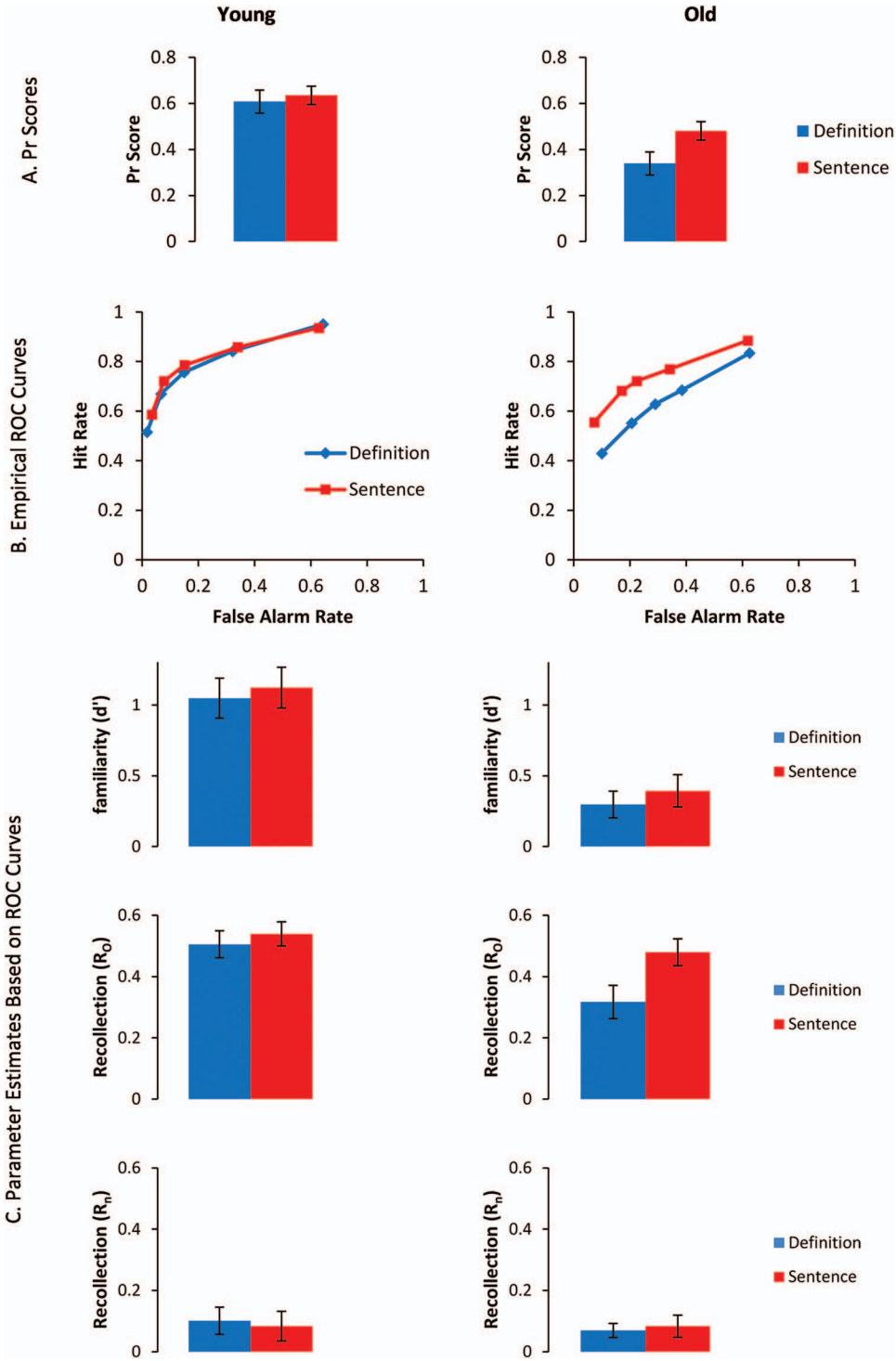

Figure 2. Behavioral Results from Experiment 1 for the young (left panels) and for the older adults (right panels). (A) Proportion of correct responses to old pairs minus incorrect responses to recombined pairs during the test phase (collapsed across *definitely old*, *probably old*, and *maybe old* responses). (B) Empirical ROC curves. For illustration purposes, the ROC curves have been generated on a group-level. (C) Parameter estimates for the contribution of familiarity and recollection to retrieval. Estimates were generated for each participant individually. Note: All error bars denote the standard error of the mean. See the online article for the color version of this figure.

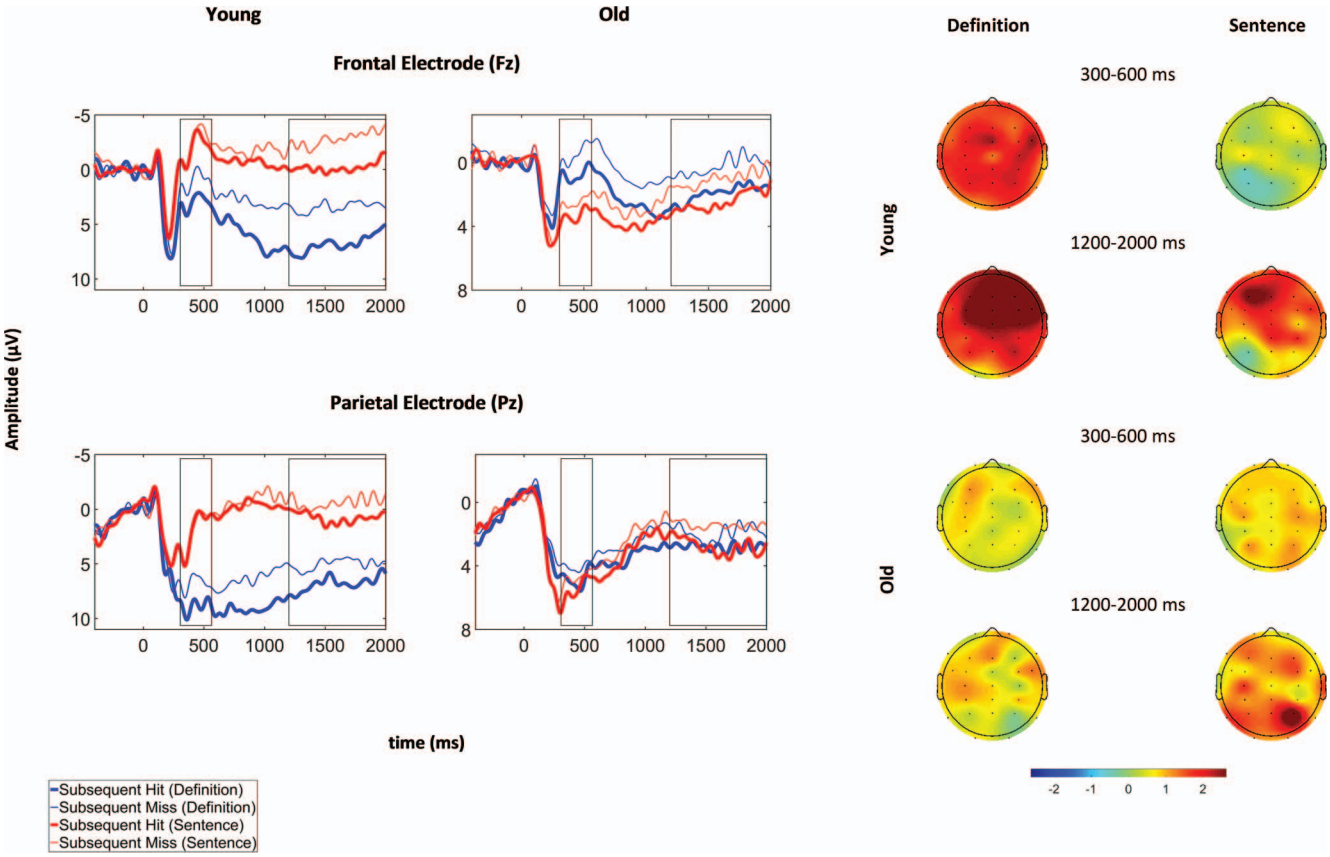

Figure 3. Grand average ERPs (a) and scalp distributions of subsequent memory effects (b) from Experiment 1. (a) Grand averages from a frontal (Fz) and a parietal (Pz) electrode for young (left panels) and older (right panels) adults, for both conditions and for trials with subsequent hits (high confidence "old" responses) and subsequent misses (medium and low confidence "old" responses and "recombined" responses). Note that amplitude scales are different for young and older adults to enhance visibility of SMEs in both conditions for the older adults. (b) Scalp distribution of the difference between subsequent hit and subsequent miss trials for the two time windows of interest, and for both age groups and conditions. See the online article for the color version of this figure.

SMEs in the older adults either. A lower-level 3 (Anteriority: frontal, central, parietal) × 3 (Laterality: left, central, right) × 2 (Subsequent Memory: subsequent hit, subsequent miss) × 2 (Condition: sentence, definition) mixed ANOVA for the older adults substantiated this impression: There was a significant SME, $F(1, 31) = 8.38$, $p = .007$, $\eta_p^2 = .21$, but there were no interactions with condition (all *p values* > .25). Finally, pairwise comparisons of the magnitude of the SME at each electrode also did not reveal any evidence for differences in the SME between conditions (all *p values* > .33).

Interim Discussion

We observed the typical reduction in associative memory performance in old adults. As we predicted, older (but not young) adults tended to differ in their recognition memory performance depending on whether word pairs were encoded in the context of an entity-defining framework allowing for unitization or together with a sentence that preserved each word as an individual unit. Strikingly, however, this pattern was in a direction reverse from our prediction: Performance was numerically lower in the definition than the sentence condition for older adults, and therefore age differences tended to actually be larger in our high compared with the low unitization condition.

There are two possible explanations for the unexpected result that older adults did not benefit from unitization encoding. First, it may not be the case that older adults can more easily retrieve unitized than nonunitized associations due to an increased reliance on their intact familiarity-based item retrieval. In support of this, an analysis of the recognition-phase ERPs of the older adults (see supplementary document and Figure S1), suggests that older adults had stronger difficulties with retrieval of the pairs in the definition condition, and that compensatory activity was presumably engaged to counteract these retrieval difficulties. However, this analysis did not provide direct evidence for or against a differential reliance on familiarity or recollection in older adults in the two conditions.

Alternatively, the unitization process itself may be more difficult for older adults, such that no robustly unitized representation is actually formed during encoding. Note that this idea is different

from a potential difference in general task difficulty between conditions in older adults. In fact, the number of "high fit" ratings given during encoding did not differ by age or condition (all p values $> .12$), speaking against a difference in general task difficulty. Rather, the idea is that specific processes needed to unitize the pairs into a coherent whole were not effectively engaged in older adults. The fact that unlike for the young adults, the encoding ERPs of older adults showed no robust difference between the two conditions somewhat supports the latter hypothesis. Before elaborating on this idea in the General Discussion, we further explore the first idea in Experiment 2.

According to the ROC analysis, the contribution of familiarity on associative memory did not differ between the two conditions, but recollection was reduced in the definition condition across both age groups. Unitization should increase reliance on familiarity-based associative retrieval and reduce reliance on recollection, so that impaired recollection in older adults should not be as critical for recognition performance. Thus, one possible reason for the fact that unitization did not alleviate the associative deficit is that the test format did not allow participants to rely sufficiently strongly on familiarity for unitization to show an effect on recognition performance. Experiment 2 tested whether an improved associative memory performance could be uncovered in the definition compared to the sentence condition in older adults with a more optimal test format.

Experiment 2

For several reasons the test format used in Experiment 1 may not have been ideal to reveal a benefit of unitization for associative memory in older adults. First, in order to distinguish old from recombined pairs, older participants in the definition condition may try to judge how plausible the pair is as a novel concept rather than relying on purely mnemonic information. There is indeed evidence that older adults tend to rely on such "plausibility strategies" in recognition tests (Reder, Wible, & Martin, 1986). However, because all recognition probes—including the recombined ones—are grammatically legal compound words and are therefore matched for plausibility, such a heuristic is of little use.

A related issue is that the new meaning of recombined pairs could, in their gist, be very similar to the concepts actually encountered during encoding. For example, if "milk taxi" (a delivery service for dairy products) was encoded, "vegetable taxi" was a possible recombined pair, which describes a quite similar concept (a delivery service for food). Older adults tend to focus on gist information during retrieval (Koutstaal & Schacter, 1997), perhaps leading to a difficulty distinguishing between old and recombined pairs. However, if this explanation was true, then an increase in false alarms in the definition condition relative to the sentence condition would be expected. Because we did not observe any differences in the false alarm rates, we can tentatively rule out the gist explanation.

Perhaps most importantly, in a standard recognition test memory traces can be retrieved based on a mixture of familiarity and recollection (Yonelinas, 2002). Unitization should result in enhanced familiarity for the entire concept. Although we assumed that older adults would rely on this familiarity signal during recognition, the test format may have biased them to rely on recollection even though it was not required. Importantly, the lower recollection parameter in the definition condition suggests that it was more challenging (especially for older adults, who exhibited generally reduced recollection relative to young adults) to recollect specifics of the studied concepts in the definition than the sentence condition, perhaps because of the focus on familiarity- and gist-based learning. In sum, older adults may not have benefitted from unitization because the characteristics of the test did not sufficiently promote familiarity-based retrieval.

Using a similar logic as Delhaye and Bastin (2018, Experiment 2), we tested this idea in the purely behavioral Experiment 2, which employed a 2AFC recognition test. This test format requires an identification of which of two alternatives (in our case an old and a recombined pair) has been previously studied and relies more strongly on familiarity than yes/no recognition (Bastin & Van der Linden, 2003). Furthermore, with this test format older adults should be less inclined to use a plausibility or gist-based recognition strategy, because intact and recombined pairs are about equal in these characteristics.

Migo, Montaldi, Norman, Quamme, and Mayes (2009) have demonstrated that familiarity contributes to recognition in 2AFC tasks particularly when the foil is similar to the target. Therefore, in our Experiment 2 an old pair was always displayed together with a recombined pair in which one of the words was identical to the corresponding old pair (e.g., "family bikini" together with "family stairs;" Figure 1C). Second, because familiarity is a faster process than recollection (Yonelinas, 2002), we attempted to prevent recollection-based responding by introducing a response deadline, which was specified based on reaction time (RT) patterns in a pilot study. In sum, since the test format encouraged familiarity-based responding and discouraged recollection or recall-to-reject, we expected that associative memory of older adults would benefit from unitization in Experiment 2.

Method

Participants. Only older adults were included, because in pilot work young adults performed at ceiling with a 2AFC test. A new sample of 44 older adults was paid €8 per hour and randomly assigned to either the definition or the sentence condition. At the end of the session they completed the same set of neuropsychological tests as in Experiment 1 (see Table 1). Participants assigned to the two conditions did not differ in age or any of the neuropsychological tests (all p *value*s $> .27$). All participants exhibited an MMSE score of at least 27.

Stimuli and procedure. The encoding and distractor phases were identical to Experiment 1. In the recognition test (Figure 1C) participants were presented with two word pairs to the left and right of the computer screen. One of these pairs was old and the other one was a recombination of one of the words from the simultaneously presented old pair with a word from a different study pair. Which of the two words from the old pair was included in the recombined pair was chosen at random, with the restriction that in half of the trials it would be the first, and in the other half it would be the second word. Participants pressed the left or the right key depending on the location of the old pair. If participants responded within 4 s, their response terminated the screen. If they did not, the word pair was replaced by the words "respond faster." Between two successive test trials a fixation cross was shown for 2 s. Eighty recognition trials were completed in pseudorandom

order: In 40 trials the old pair was on the left, and in 40 trials it was on the right side of the screen. A break was allowed after each set of 40 trials. Before the recognition phase, participants completed three practice trials; the practice could be repeated if necessary.

Data analysis. We calculated the percent of correct recognition decisions separately for "high fit" and "low fit" pairs as well as response times in the recognition test and the percentage of trials in which participants failed to provide a response within 4 s ("time outs"). Because one participant in the sentence condition had no "low fit" trials, the data from this participant were excluded from the analysis. We also analyzed whether the relative effect of the two encoding conditions on recognition performance differed for the older adults between Experiments 1 and 2. The dependent variables were the hit rates from Experiment 1 and percent correct recognition decisions for Experiment 2. Due to the difference in the performance measure, main effects for experiment have to be interpreted with caution.

Results

Participants in the two groups did not significantly differ in the accuracy of detecting the old pair (see Table 2): A 2 (Condition) × 2 (Fit Rating) ANOVA revealed no significant main effects or interactions (all p *values*>.19). A 2 (Experiment) × 2 (Condition) × 2 (Fit Rating) ANOVA contrasting performance (hit rates and percent correct, respectively) of the older adults in Experiments 1 and 2 revealed a significant condition by experiment interaction, $F(1, 81) = 4.06$, $p < .05$, $\eta_p^2 = .05$: Recognition performance differed between the sentence and definition condition in Experiment 1, but not in Experiment 2. In addition, the ANOVA revealed main effects for fit rating, $F(1, 81) = 6.56$, $p = .01$, $\eta_p^2 = .08$, and condition, $F(1, 81) = 4.72$, $p = .03$, $\eta_p^2 = .06$, qualified by an interaction of fit rating and condition, $F(1, 81) = 4.52$, $p = .04$, $\eta_p^2 = .05$. In the definition condition the participants' fit judgment had a stronger influence on recognition performance than in the sentence condition. Finally, a fit rating by experiment interaction, $F(1, 81) = 4.29$, $p = .04$, $\eta_p^2 = .05$, reflected the fact that in Experiment 1 performance dropped more strongly from high to low fit trials than in Experiment 2.

An independent samples t test revealed that participants in the sentence condition ($M = 5.15$) showed a trend for a higher number of time out trials compared with the definition condition ($M = 2.77$), $t(42) = 1.77$, $p = .08$, $d = .54$.

A 2 (Fit Rating) × 2 (Accuracy) × 2 (Condition) ANOVA on the mean RTs during the recognition test (see Figure 4) resulted in a main effect for accuracy, $F(1, 41) = 5.13$, $p = .03$, $\eta_p^2 = .11$: Correct responses were faster than incorrect responses. Due to a significant fit rating by condition interaction, $F(1, 41) = 7.92$, $p = .007$, $\eta_p^2 = .16$, we conducted separate ANOVAs for "high fit" and "low fit" trials.

For low-fit trials, a 2 (Accuracy) × 2 (Condition) ANOVA revealed no main effects and no interaction (all p *values* > .59). For high fit trials, the analogous ANOVA revealed a main effect for accuracy, $F(1, 42) = 10.39$, $p = .002$, $\eta_p^2 = .20$. The trends for a main effect for condition, $F(1, 42) = 3.85$, $p = .06$, and an interaction, $F(1, 42) = 3.490$, $p = .07$, were not significant.

Interim Discussion

By contrast to a standard recognition task (Experiment 1), in a 2AFC test (Experiment 2) performance of older participants in the definition and sentence condition did not differ. Therefore, consistently with our prediction associative memory benefited more from unitization in Experiment 2 than Experiment 1, presumably due to a larger contribution of familiarity-based remembering. However, although several measures were taken to maximize the contribution of familiarity to retrieval, no advantage for the definition condition emerged.

Interestingly, using a similar encoding task to ours with preexisting versus nonexisting compound words, Delhaye and Bastin (2018) found that with a 2AFC task, the associative deficit of older adults was entirely abolished in both encoding conditions. In our Experiment 2, older adults still exhibited lower associative memory performance than young adults. However, our findings align with those of Delhaye and Bastin (2018) in that, contrary to both their expectation and ours, the 2AFC task did not lead to a *disproportionate* reduction of the associative memory deficit of older adults in the high-compared to the low unitization condition. Potential reasons accounting for the fact that we did not find a reduction of the associative memory deficit in the definition condition are discussed further in the general discussion.

We found a tendency for recognition responses to be faster in the definition than the sentence condition, as measured by RTs (for the high fit trials) and number of time outs. However, these effects were statistically not significant, and they did not translate into improved memory accuracy in the definition condition.

General Discussion

We investigated how varying the degree to which word pairs are unitized into a coherent, new concept affects the success and manner in which young and older adults encode and retrieve novel associations. In a standard associative recognition test, associative memory of older adults did not benefit from the definition condition (which encourages unitization). Rather, the associative memory impairment tended to be more pronounced in this condition,

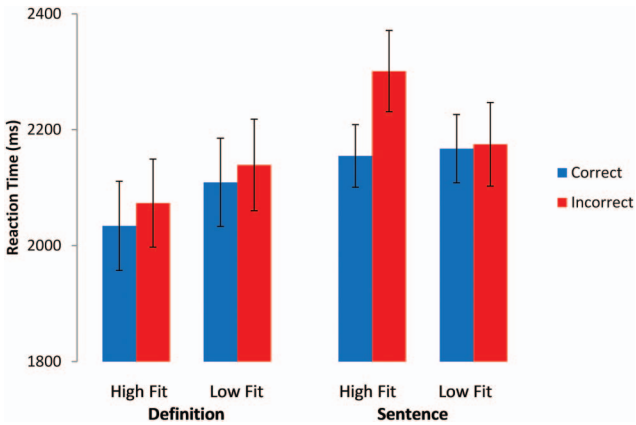

Figure 4. Mean RTs in the recognition test for participants in the definition and the sentence groups, by "fit" rating and response accuracy. Error bars represent the standard error of the mean. See the online article for the color version of this figure.

compared with the sentence (or control) condition (Experiment 1). ROC analyses suggested that the recognition test may not have relied on familiarity enough to reveal the expected effect of unitization on the associative deficit. With a 2AFC test designed to maximize the contribution of familiarity to associative recognition, performance differences between conditions were abolished in older adults, but still no memory advantage for the unitization condition emerged (Experiment 2).

In the first part of the discussion, we will debate on the basis of the behavioral result patterns whether unitization contributed to associative retrieval in older adults. Because the evidence is rather inconclusive, it appears that the older adults had at least some difficulty with unitization already during encoding. In the second part, we therefore consult the ERP patterns for potential age-related neurocognitive differences of encoding in the two conditions. In the last part, we debate the boundary conditions to exhaust the full potential of unitization as an encoding strategy to improve older adults' associative memory.

Did Unitization Lead to Familiarity-Based Retrieval of Associations?

In line with other studies (e.g., Bader et al., 2010), young adults' total recognition performance did not differ between the definition and the sentence condition (Experiment 1). It is therefore important to keep in mind that unitization does not necessarily lead to improved memory performance per se. Rather, unitization changes what kind of memory signals are used to distinguish old from new pairs, that is, it changes the *manner* in which the pairs are retrieved. Evidence from prior studies, including ERP data from the recognition phase of the present dataset (Kamp et al., 2016) supports the idea that retrieval of old pairs in the definition condition relies relatively more strongly on familiarity than recollection, compared with the sentence condition in young adults. This pattern is supported by the ROC analysis of the present study: The familiarity parameter did not differ between conditions, while recollection was reduced in the definition condition. It must be noted that the *absolute magnitude* of the familiarity signal was not enhanced in the definition condition, neither as reflected by ERP measures (Kamp et al., 2016), nor as reflected by ROC parameters of familiarity. Rather, the *relative* contribution of familiarity to associative retrieval was larger in the definition than the sentence condition.

This pattern was found in both age groups and provides some indication that unitization of the word pair may have been successful in older adults. However, the fact that the pattern was driven by reduced recollection parameters rather than increased familiarity parameters somewhat complicates the interpretation. In Experiment 2, a test format that relies more strongly on familiarity-based retrieval led to a relative performance improvement in the definition condition compared to Experiment 1. Nevertheless, in the light of evidence that recollection (which is required for retrieval of nonunitized associations) is more strongly affected by aging than familiarity and item memory (which can support retrieval of unitized associations; Old & Naveh-Benjamin, 2008), the fact that the definition condition did not lead to an absolute performance advantage, even in the more familiarity-reliant Experiment 2 is difficult to reconcile with the idea that unitization was completely successful.

Notably, in both age groups the extent to which the participants indicated that they could imagine the new concept/sentence well had a stronger effect on associative recognition performance in the definition than the sentence condition. This supports the idea that unitization had an impact on associative memory: Formation of a coherent new concept requires the ability to successfully integrate the components of the association, but for inter-item relational encoding, the quality of the integration into a coherent whole is not critical. However, in Experiment 2, the condition difference in this effect was much smaller and did not reach significance. Taken together, although there are some hints for an effect of unitization on associative memory in both age groups, the evidence for older adults is not conclusive. This raises the question whether older adults had difficulties with the unitization process already during encoding.

Encoding Mechanisms of Unitization in Young and Older Adults

Main effects for condition on ERP amplitudes. In young adults, word pairs in the definition condition elicited more positive-going waveforms for the definition than the sentence condition beginning in the early time window and extending over the entire segment. This may reflect an increased P300 (see Donchin, 1981) and/or a decreased N400 (see Kutas & Federmeier, 2011). Alternatively, there may have been a contribution from the frontal positivity that has been reported to index the encounter of linguistic information that is unexpected but plausible given a preceding context (Federmeier, Kutas, & Schul, 2010). Regardless of the precise componential identity of this effect, based on the functional significance of each possible ERP component it may reflect, it is highly likely that it reflects the processing of new information, which is unexpected but plausible in the light of a preceding context. Our interpretation of the main effect for condition, therefore, is that upon encounter in the definition (but not the sentence), young adults generated expectations about what word pairs may follow. Young adults therefore appear to adopt a "proactive" unitization encoding strategy by engaging resources in deriving predictions regarding the upcoming word pair before it is presented, which in turn facilitates the integration into a new holistic concept when the word pair is encountered. Because the sentences are less constraining as to what kind of word pair will follow, predictive processing is not as useful in this condition. This results in the smaller overall expectancy mismatch and an attenuated positivity elicited by the onset of the word pair in the sentence compared with the definition condition.

In the older adults, the only difference between the conditions in overall ERP amplitudes in the early time window was a larger anterior negativity, peaking 400 ms–600 ms after word pair onset, in the definition than the sentence condition. Importantly, the distribution and polarity of this effect are different from the condition effects for the young adults. This frontal negativity may bear some relation to the N400-like activity recorded from intracranial electrodes within anterior MTL (McCarthy, Nobre, Bentin, & Spencer, 1995) and may reflect semantic integration processes. Indeed, the fact that word pairs in the definition condition elicited a larger negativity in older adults than the sentence condition suggests that it was difficult for the older adults to integrate the word pairs into the context of the definition, perhaps due to lacking

or impoverished predictive processing and expectancy build up after encountering the definition. Taken together, the ERPs elicited by the onset of the word pair are consistent with the idea that older adults in the unitization condition did not engage in a proactive encoding strategy similar to the young adults. This idea is in line with prior evidence that older adults do not utilize information on cues preceding a target to the same extent as young adults in linguistic (Federmeier et al., 2010) and executive control tasks (e.g., Schmitt, Ferdinand, & Kray, 2014). Our finding therefore suggests a reduction of frontal-lobe-reliant top-down control in older adults during unitization tasks.

One way to more directly test the idea that young and older adults differ in the extent to which they adopt proactive encoding strategies is to examine prestimulus ERPs and SMEs. Prestimulus SMEs are typically calculated to the onset of a prestimulus task or modality cue (e.g., Otten, Quayle, & Puvaneswaran, 2010). In the case of our experiment, prestimulus ERPs and SMEs would have to be computed with reference to the onset of the definition or sentence. However, the definitions and sentences were not matched for visual and semantic features and also varied among several dimensions within each condition, complicating this kind of analysis. Future studies should use a design that allows for a prestimulus ERP/SME analysis with a similar encoding task to ours to examine this issue.

SMEs. For the young adults (Kamp et al., 2017), in the early time window an SME was observed only in the definition condition, suggesting that it indexes a process that is engaged when the word pair is unitized into a coherent concept, but not in an inter-item encoding condition. The overlap between the early SME and the enhanced positivity in the definition condition suggest that the facilitated integration into the new concept due to predictive processing is directly relevant for successful item or unitization encoding. The later frontal SMEs did not differ between the conditions. Thus, young adults engage at least partially different encoding mechanisms in unitization, compared to inter-item encoding tasks.

In older adults, ERPs exhibited more positive-going amplitudes for subsequent hits versus misses in both time windows. There was no evidence for differences in SMEs between the two conditions, indicating that older adults did not engage differential encoding mechanisms in the two conditions. Under the assumption that the early SME reflects item encoding while the later SME reflects encoding of associations (Kamp et al., 2017), this suggests that item and inter-item encoding mechanisms nondistinctly contributed to encoding in both conditions.

Also noteworthy is that there was no evidence that any of the SMEs differed between scalp locations in older adults. Our results are in line with the dedifferentiation hypothesis (Li, Lindenberger, & Sikström, 2001), which proposes that neural signals become more "noisy" with aging and consequently the corresponding representations become less distinct. In our task, older adults may be unable to flexibly adjust and optimize the neurocognitive mechanisms of encoding when a selective benefit from an increase in either item or inter-item associative encoding can be expected.

Unitization in Aging

In contrast to some prior studies (Ahmad & Hockley, 2014; Bastin et al., 2013; Zheng et al., 2015), but in line with others (Bridger et al., 2017; Delhaye & Bastin, 2018; Jäger et al., 2010) we did not uncover an associative memory benefit of older adults in an encoding task that is known—at least in young adults—to promote unitization. In Experiment 1 the age difference was in fact larger in the unitization than the control condition. This result is at first glance in conflict with Badham, Estes, and Maylor (2012), who instructed participants to encode word pairs that were either unrelated, semantically related, or linked through an integrative relation. Pairs in the latter condition were similar to our stimuli in that they formed a novel compound word, although no definition was provided to help in forming the concept and no explicit instructions were given to do so. The authors found that the age-related deficit in a cued-recall test was reduced both for semantically related and integrative, compared to unrelated word pairs. However, as the authors note, cueing recall with a word from an integrative pair (as compared with an unrelated pair) may specifically facilitate *search* for the second word, a process that is not relevant for associative recognition. Therefore, a direct comparison of these results with ours is difficult.

One possibility why encoding in the definition condition was difficult for older adults is that novel concepts were not supported by, and perhaps even in conflict with already existing world knowledge. Although the task of the present paradigm was designed to provide ample external support to promote unitization by means of the definition, such a conflict of the new concepts with preexisting knowledge could have led to a strong need for controlled top-down processing. Encoding tasks with such demands have been shown to increase, rather than decrease, age differences in episodic memory (Luo et al., 2007). Therefore, similarly to Bridger et al. (2017), unitization in the present paradigm may not take advantage of preexisting semantic knowledge enough for older adults to boost item encoding and overall recognition memory.

The ERP results suggest that unlike the young adults, older adults (a) did not use a "proactive" encoding strategy and (b) did not engage in differential encoding mechanisms in the two conditions. An open question is whether older adults benefit from instructions or training to use a proactive encoding strategy and/or to engage distinct encoding mechanisms in the definition condition, thereby making the circumstances of encoding more similar to the young adults. Such instructions would perhaps encourage older adults to more actively engage in generating the new concept, perhaps resulting in a facilitative effect of unitization on associative memory in older adults. Another advantageous modification to our paradigm could be to increase the personal relevance of the stimuli to the participants by, for example, asking them to provide a rating of how pleasant they personally find the new concept.

Finally, an important implication from our study is that the effects of encoding strategies that promote unitization on recognition performance strongly depend on the manner in which recognition is tested. Memory tests that rely strongly on recollection do not provide the ideal conditions for unitization to alleviate the associative memory deficit in aging, and this factor appears to be especially essential when unitization encoding is only partially successful.

References

Ahmad, F. N., Fernandes, M., & Hockley, W. E. (2015). Improving associative memory in older adults with unitization. *Aging, Neuropsychology and Cognition, 22,* 452–472. http://dx.doi.org/10.1080/13825585.2014.980216

Ahmad, F. N., & Hockley, W. E. (2014). The role of familiarity in associative recognition of unitized compound word pairs. *Quarterly Journal of Experimental Psychology, 67,* 2301–2324. http://dx.doi.org/10.1080/17470218.2014.923007

Bader, R., Mecklinger, A., Hoppstädter, M., & Meyer, P. (2010). Recognition memory for one-trial-unitized word pairs: Evidence from event-related potentials. *NeuroImage, 50,* 772–781. http://dx.doi.org/10.1016/j.neuroimage.2009.12.100

Bader, R., Opitz, B., Reith, W., & Mecklinger, A. (2014). Is a novel conceptual unit more than the sum of its parts?: FMRI evidence from an associative recognition memory study. *Neuropsychologia, 61,* 123–134. http://dx.doi.org/10.1016/j.neuropsychologia.2014.06.006

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

Bastin, C., Diana, R. A., Simon, J., Collette, F., Yonelinas, A. P., & Salmon, E. (2013). Associative memory in aging: The effect of unitization on source memory. *Psychology and Aging, 28,* 275–283. http://dx.doi.org/10.1037/a0031566

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology, 17,* 14–24. http://dx.doi.org/10.1037/0894-4105.17.1.14

Bridger, E. K., Kursawe, A. L., Bader, R., Tibon, R., Gronau, N., Levy, D. A., & Mecklinger, A. (2017). Age effects on associative memory for novel picture pairings. *Brain Research, 1664,* 102–115. http://dx.doi.org/10.1016/j.brainres.2017.03.031

Cansino, S., Trejo-Morales, P., & Hernández-Ramos, E. (2010). Age-related changes in neural activity during source memory encoding in young, middle-aged and elderly adults. *Neuropsychologia, 48,* 2537–2549. http://dx.doi.org/10.1016/j.neuropsychologia.2010.04.032

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769–786. http://dx.doi.org/10.3758/BF03196772

Craik, F. I. M., & Rose, N. S. (2012). Memory encoding and aging: A neurocognitive perspective. *Neuroscience and Biobehavioral Reviews, 36,* 1729–1739. http://dx.doi.org/10.1016/j.neubiorev.2011.11.007

Delhaye, E., & Bastin, C. (2018). The impact of aging on associative memory for preexisting unitized associations. *Aging, Neuropsychology and Cognition, 25,* 70–98. http://dx.doi.org/10.1080/13825585.2016.1263725

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). The effects of unitization on familiarity-based source memory: Testing a behavioral prediction derived from neuroimaging data. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 730–740. http://dx.doi.org/10.1037/0278-7393.34.4.730

Dodson, C. S., Prinzmetal, W., & Shimamura, A. P. (1998). Using Excel to estimate parameters from observed data: An example from source memory data. *Behavior Research Methods, Instruments, & Computers, 30,* 517–526. http://dx.doi.org/10.3758/BF03200685

Donchin, E. (1981). Presidential address, 1980. Surprise!. Surprise? *Psychophysiology, 18,* 493–513. http://dx.doi.org/10.1111/j.1469-8986.1981.tb01815.x

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language, 115,* 149–161. http://dx.doi.org/10.1016/j.bandl.2010.07.006

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

Friedman, D., Ritter, W., & Snodgrass, J. G. (1996). ERPs during study as a function of subsequent direct and indirect memory testing in young and old adults. *Cognitive Brain Research, 4,* 1–13. http://dx.doi.org/10.1016/0926-6410(95)00041-0

Friedman, D., & Trott, C. (2000). An event-related potential study of encoding in young and older adults. *Neuropsychologia, 38,* 542–557. http://dx.doi.org/10.1016/S0028-3932(99)00122-0

Graf, P., & Schacter, D. L. (1989). Unitization and grouping mediate dissociations in memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 930–940. http://dx.doi.org/10.1037/0278-7393.15.5.930

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55,* 468–484. http://dx.doi.org/10.1016/0013-4694(83)90135-9

Gutchess, A. H., Ieuji, Y., & Federmeier, K. D. (2007). Event-related potentials reveal age differences in the encoding and recognition of scenes. *Journal of Cognitive Neuroscience, 19,* 1089–1103. http://dx.doi.org/10.1162/jocn.2007.19.7.1089

Haskins, A. L., Yonelinas, A. P., Quamme, J. R., & Ranganath, C. (2008). Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. *Neuron, 59,* 554–560. http://dx.doi.org/10.1016/j.neuron.2008.07.035

Healy, M. R., Light, L. L., & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 768–788. http://dx.doi.org/10.1037/0278-7393.31.4.768

Jäger, T., & Mecklinger, A. (2009). Familiarity supports associative recognition memory for face stimuli that can be unitised: Evidence from receiver operating characteristics. *The European Journal of Cognitive Psychology, 21,* 35–60. http://dx.doi.org/10.1080/09541440802003140

Jäger, T., Mecklinger, A., & Kipp, K. H. (2006). Intra- and inter-item associations doubly dissociate the electrophysiological correlates of familiarity and recollection. *Neuron, 52,* 535–545. http://dx.doi.org/10.1016/j.neuron.2006.09.013

Jäger, T., Mecklinger, A., & Kliegel, M. (2010). Associative recognition memory for faces: More pronounced age-related impairments in binding intra-than inter-item associations. *Experimental Aging Research, 36,* 123–139. http://dx.doi.org/10.1080/03610731003613391

Kamp, S.-M., Bader, R., & Mecklinger, A. (2016). The effect of unitizing word pairs on recollection- vs. familiarity-based retrieval—Further evidence from ERPs. *Advances in Cognitive Psychology, 12,* 169–178. http://dx.doi.org/10.5709/acp-0196-2

Kamp, S.-M., Bader, R., & Mecklinger, A. (2017). ERP subsequent memory effects differ between inter-item and unitization encoding tasks. *Frontiers in Human Neuroscience, 11,* 30. http://dx.doi.org/10.3389/fnhum.2017.00030

Kamp, S.-M., & Zimmer, H. D. (2015). Contributions of attention and elaboration to associative encoding in young and older adults. *Neuropsychologia, 75,* 252–264. http://dx.doi.org/10.1016/j.neuropsychologia.2015.06.026

Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology, 16,* 177–216. http://dx.doi.org/10.1016/0010-0285(84)90007-0

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

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,* 621–647. http://dx.doi.org/10.1146/annurev.psych.093008.131123

Li, J., Morcom, A. M., & Rugg, M. D. (2004). The effects of age on the neural correlates of successful episodic retrieval: An ERP study. *Cognitive, Affective & Behavioral Neuroscience, 4,* 279–293. http://dx.doi.org/10.3758/CABN.4.3.279

Li, S.-C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences, 5,* 479–486. http://dx.doi.org/10.1016/S1364-6613(00)01769-1

Light, L. L., Prull, M. W., La Voie, D. J., & Healy, M. R. (2000). Dual-process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 238–300). Oxford, UK: Oxford University Press.

Lindenberger, U., Mayr, U., & Kliegl, R. (1993). Speed and intelligence in old age. *Psychology and Aging, 8,* 207–220. http://dx.doi.org/10.1037/0882-7974.8.2.207

Luo, L., Hendriks, T., & Craik, F. I. (2007). Age differences in recollection: Three patterns of enhanced encoding. *Psychology and Aging, 22,* 269–280. http://dx.doi.org/10.1037/0882-7974.22.2.269

McCarthy, G., Nobre, A. C., Bentin, S., & Spencer, D. D. (1995). Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *The Journal of Neuroscience, 15,* 1080–1089.

Migo, E., Montaldi, D., Norman, K. A., Quamme, J., & Mayes, A. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 62,* 1198–1215. http://dx.doi.org/10.1080/17470210802391599

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging, 22,* 202–208. http://dx.doi.org/10.1037/0882-7974.22.1.202

Ofen, N., & Shing, Y. L. (2013). From perception to memory: Changes in memory systems across the lifespan. *Neuroscience and Biobehavioral Reviews, 37,* 2258–2267. http://dx.doi.org/10.1016/j.neubiorev.2013.04.006

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23,* 104–118. http://dx.doi.org/10.1037/0882-7974.23.1.104

Otten, L. J., Quayle, A. H., & Puvaneswaran, B. (2010). Prestimulus subsequent memory effects for auditory and visual events. *Journal of Cognitive Neuroscience, 22,* 1212–1223. http://dx.doi.org/10.1162/jocn.2009.21298

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences, 6,* 93–102. http://dx.doi.org/10.1016/S1364-6613(00)01845-3

Parks, C. M., & Yonelinas, A. P. (2015). The importance of unitization for familiarity-based learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 881–903. http://dx.doi.org/10.1037/xlm0000068

Quamme, J. R., Yonelinas, A. P., & Norman, K. A. (2007). Effect of unitization on associative recognition in amnesia. *Hippocampus, 17,* 192–200. http://dx.doi.org/10.1002/hipo.20257

Reder, L. M., Wible, C., & Martin, J. (1986). Differential memory changes with age: Exact retrieval versus plausible inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 72–81. http://dx.doi.org/10.1037/0278-7393.12.1.72

Rhodes, S. M., & Donaldson, D. I. (2008). Electrophysiological evidence for the effect of interactive imagery on episodic memory: Encouraging familiarity for non-unitized stimuli during associative recognition. *NeuroImage, 39,* 873–884. http://dx.doi.org/10.1016/j.neuroimage.2007.08.041

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11,* 251–257. http://dx.doi.org/10.1016/j.tics.2007.04.004

Scheuplein, A.-L., Bridger, E. K., & Mecklinger, A. (2014). Is faster better? Effects of response deadline on ERP correlates of recognition memory in younger and older adults. *Brain Research, 1582,* 139–153. http://dx.doi.org/10.1016/j.brainres.2014.07.025

Schmitt, H., Ferdinand, N. K., & Kray, J. (2014). Age-differential effects on updating cue information: Evidence from event-related potentials. *Cognitive, Affective & Behavioral Neuroscience, 14,* 1115–1131. http://dx.doi.org/10.3758/s13415-014-0268-9

Shing, Y. L., Werkle-Bergner, M., Li, S. C., & Lindenberger, U. (2008). Associative and strategic components of episodic memory: A life-span dissociation. *Journal of Experimental Psychology: General, 137,* 495–513. http://dx.doi.org/10.1037/0096-3445.137.3.495

Tibon, R., Gronau, N., Scheuplein, A. L., Mecklinger, A., & Levy, D. A. (2014). Associative recognition processes are modulated by the semantic unitizability of memoranda. *Brain and Cognition, 92C,* 19–31. http://dx.doi.org/10.1016/j.bandc.2014.09.009

Wechsler, D. (2008). *Wechsler adult intelligence scale–fourth edition (WAIS–IV).* San Antonio, TX: NCS Pearson.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354. http://dx.doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517. http://dx.doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., & Soltani, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review, 6,* 654–661. http://dx.doi.org/10.3758/BF03212975

Zheng, Z., Li, J., Xiao, F., Broster, L. S., & Jiang, Y. (2015). Electrophysiological evidence for the effects of unitization on associative recognition memory in older adults. *Neurobiology of Learning and Memory, 121,* 59–71. http://dx.doi.org/10.1016/j.nlm.2015.03.006

Received September 15, 2017
Revision received February 26, 2018
Accepted March 4, 2018 ■