A Specificity Principle of Memory: Evidence From Aging and Associative Memory

Nathaniel R. Greene and Moshe Naveh-Benjamin
Department of Psychological Sciences, University of Missouri

Abstract
The ability to remember associations among components of an event, which is central to episodic memory, declines with normal aging. In accord with the specificity principle of memory, these declines may occur because associative memory requires retrieval of specific information. Guided by this principle, we endeavored to determine whether ubiquitous age-related deficits in associative memory are restricted to specific representations or extend to the gist of associations. Young and older adults (30 each in Experiment 1, 40 each in Experiment 2) studied face–scene pairs and then performed associative-recognition tests following variable delays. Whereas both young and older adults could retrieve the gist of associations, older adults were impaired in their ability to retrieve more specific representations. Our results also show that associations can be retrieved from multiple levels of specificity, suggesting that episodic memory might be accessed on a continuum of specificity.

Scientific principles describe invariants of natural phenomena, but efforts to identify principles of memory have been scarce and often unsuccessful (Roediger, 2008). However, Surprenant and Neath (2009) provided compelling evidence that several principles may exist across all memory systems. Since their book was published, there have been relatively few attempts to test their proposed principles and fewer attempts to assess whether these principles can explain changes in memory performance that occur with normal aging. Here, we report on their suggested specificity principle, according to which vulnerability to forgetting is highest for tasks requiring individuals to retrieve specific information.

The specificity principle may provide insight into causes of pervasive age-related deficits in episodic memory, particularly for the associations among components of an episode (e.g., Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008). Episodic memories are bounded representations of contextual, temporal, and spatial components (Tulving, 1983). According to the associative-deficit hypothesis (Naveh-Benjamin, 2000), older adults can remember individual components of

an episode relatively well, compared with young adults, but often fail to encode or retrieve the associations among these components. These failures contribute to memory errors, such as misremembering where a person was previously encountered (Chen & Naveh-Benjamin, 2012), and can explain why older adults often forget the source of information (Boywitt, Kuhlmann, & Meiser, 2012). We propose that at least some of these deficits arise because associative-memory tasks require individuals to retrieve specific information, a process impaired by normal aging.

Traditionally, associative-memory tasks require individuals to distinguish previously studied associations (A–B and C–D) from recombined associations (A–C or

Corresponding Authors:
Nathaniel R. Greene, University of Missouri, Department of Psychological Sciences, 9J McAlester Hall, Columbia, MO 65211
E-mail: ngwfm@mail.missouri.edu

Moshe Naveh-Benjamin, University of Missouri, Department of Psychological Sciences, 106 McAlester Hall, Columbia, MO 65211
E-mail: navehbenjaminm@missouri.edu

B–D). To be successful at this task, one must retrieve some specific information about which items were associated at encoding. Because older adults are prone to associative-memory errors, an important unanswered question remains: How much specific information about associations can older adults retrieve? That is, do they fail to retrieve specific bounded representations, or do these deficits also extend to retrieval of the gist of associations?

The distinction between specific and gist memory is central to fuzzy-trace theory (Reyna & Brainerd, 1995), which postulates that the surface-level details and the meaning of an episode are processed in parallel in a verbatim and gist trace, respectively. Fuzzy-trace theory is commonly used to explain false-memory phenomena, such as those observed in the Deese-Roediger-McDermott paradigm (Roediger & McDermott, 1995), in which individuals study a list of items (e.g., "dream," "rest," and "bed") that are closely related to an unpresented lure ("sleep"). False recognition of the lure is well documented in older adults, which is taken as evidence that they use a gist-based processing strategy (for a recent review, see Devitt & Schacter, 2016).

Falsely recognizing a lure in this paradigm suggests that older adults remember the semantic gist of the studied material (e.g., "dream," "rest," and "bed" are related to "sleep"), but it does not necessarily indicate that they retrieve a fuzzier episodic representation of a particular event. A more direct way to measure this is to assess specific versus gist memory for the associations among components of an episode that lie at the core of episodic memory.

There have been a few attempts to assess whether older adults can remember the gist of associations, such as studies showing that they can retrieve the gist of item–price pairs (Castel, 2005; Castel, McGillivray, & Worden, 2013). However, it is possible that older adults could remember the gist of what price was associated with a given item in those studies because they could rely on preexisting knowledge (e.g., "A gallon of milk is usually $2 to $3"). This may not be possible for other associations, such as face–scene pairs. Although it is conceivable that older adults could use a preexisting schema to remember the gist of where they saw a face (e.g., "This old man was in some park"), it is also possible that alternative schemas might interfere with this ability (e.g., "There may have been a young woman in this park"), which might result in failures to retrieve the gist of an association.

Further, it has not been determined whether young or older adults can retrieve associations with varying amounts of specificity. For example, memory of an old man in a park might be retrieved with high specificity ("He was in *this* park"), with less specificity ("He was

in *some* park"), or with only a fuzzy representation reinstated ("He was somewhere in nature"). Although fuzzy-trace theory alludes to the possibility of hierarchies of gist (Brainerd & Reyna, 2015; Reyna & Brainerd, 1991), these hierarchies mostly pertain to semantic representations. In contrast, Craik (2002) proposed that memory might exist on a continuum with no categorical break between episodic and semantic representations. By this account, there could be multiple levels of specificity for an individual episode. We conducted the first test of this proposal in the context of associative memory.

The Present Study

In two experiments, we tested the hypothesis that age-related associative-memory deficits result, in part, from inabilities in retrieving specific representations but that older adults could retrieve these associations at less specific levels of representation. We used face–scene pairs, which simulate remembering where someone was encountered (Gruppuso, Lindsay, & Masson, 2007), because older adults have demonstrated an associative deficit for face–scene pairs in both short- and long-term memory (Chen & Naveh-Benjamin, 2012). To measure specific and gist memory, we used the simplified-conjoint-recognition (SCR) paradigm and its multinomial-processing-tree (MPT) model (Stahl & Klauer, 2008). Although the paradigm is based on fuzzy-trace theory, in Experiment 2, we modified it to account for the possibility of different levels of specificity. Finally, principles of memory are proposed to operate on all time scales (Surprenant & Neath, 2009), so we measured specific and gist memory for associations following various retention intervals between encoding and retrieval.

Experiment 1

Method

Participants. Thirty young adults (60% female, age: range = 18–23 years, *M* = 19.00, *SD* = 1.14) participated in exchange for psychology research credits. Thirty older adult participants (70% female; age: range = 62–80 years, *M* = 72.83, *SD* = 4.39) were recruited from the surrounding community and were given $15 in compensation for their time and travel expenses. All older adults reported being in good health, had normal or corrected-to-normal hearing and vision, and scored 27 or higher on the Mini-Mental State Exam (Folstein, Folstein, & McHugh, 1975). Because analyses were based on Bayesian posterior estimation, a formal power analysis, which is non-Bayesian by nature, was not conducted, given that evidence for a null effect is equally informative in the Bayesian framework (e.g., Kruschke, 2011, 2018). Instead, we elected to

use a sample size of 30 participants per age group, as this compares favorably with the sample sizes in the original SCR paradigm (Stahl & Klauer, 2008). On reaching this sample size, we used a Bayesian-prior sensitivity analysis, specifying our MPT models with priors that varied across three levels of informativeness. This type of analysis tests whether the results are robust to choice of prior. In our case, they were (see https://osf.io/fpcmu/). It is important to note that optional stopping (e.g., terminating data collection at the elected sample size) does not bias the obtained results in the Bayesian framework (Rouder, 2014).

Stimuli. We paired 168 faces with 168 scenes. Faces were drawn from the FACES database (Ebner, Riediger, & Lindenberger, 2010), with 42 faces from each of four age- and gender-normed categories (i.e., young female, young male, old female, old male). Scenes were drawn from a categorized scene pool (Konkle, Brady, Alvarez, & Oliva, 2010), with two exemplars from each of 84 categories (e.g., two parks, two malls).

We divided 144 face–scene pairs evenly across 12 experimental blocks, and the remaining 24 pairs were presented in two practice blocks. Each block featured six face–scene pair categories in which both exemplars of a given scene category were paired with faces from the same age and gender category. For example, two parks were paired with two old male faces in one study block. This manipulation was designed to enhance schematic support for a given association (e.g., "Parks are paired with old men") to potentially reduce the influence of other preexisting schemas (e.g., "Young men, young women, or old women could also be in parks"). Participants thus needed to maintain 12 unique associations (e.g., Old Man 1–Park 1, Old Man 2–Park 2) but only 6 gist-based categorized associations (e.g., old man–park).

Pairs were automated using E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2012) and presented on a 20-in. flat-screen LED monitor (ASUS, Beitou District, Taipei, Taiwan) with a resolution of 1,920 × 1,080 pixels. Faces appeared to the left of the center of the screen and scenes to the right. We also constructed a list of 136 populous United States cities and their corresponding states for a verbal semantic-memory filler task on blocks featuring a longer delay between the study and test phases.

Procedure. All participants provided informed consent prior to participation. Participants were tested individually with the experimenter present. The experimental session lasted about an hour and was divided into 2 practice and 12 experimental blocks. Each block consisted of a study phase and a test phase (see Fig. 1a). During the study phase, face–scene pairs were presented on screen sequentially at a rate of one pair every 4 s, with a 0.5-s interpair interval. Participants were instructed to study the pairs for a forthcoming memory test. They were informed that no two pairs would be identical but that some pairs might appear to be quite similar. Twelve face–scene pairs from the set of all possible pairs were pseudorandomly intermixed during a given study block, with the only constraint being that both pairs from a given categorized pairing had to appear in the same block. In 6 of the blocks, the test phase followed approximately 5 s to 8 s after the last pair disappeared from the screen (short delay), whereas in the remaining 6 blocks, the test phase occurred approximately 2.5 min following a verbal semantic-memory task (long delay).

In the short-delay blocks, the word "Wait" appeared, centered on screen for 5 s between the study and test phases and was followed by a test prompt informing participants that the test phase was about to begin. Participants pressed the space bar to initiate the test phase, or it began automatically after 3 s. In the long-delay blocks, an instruction prompt for the filler activity appeared after the end of the study phase. Participants initiated the task by pressing the space bar, or it began automatically after 5 s. For the semantic-memory task, the names of 21 United States cities appeared on screen sequentially, one at a time, for 5 s, and participants were instructed to say aloud the state in which each city was located. The correct answer was shown for an additional 2 s. At the end of the task, a test prompt appeared as in the short-delay blocks and was terminated by pressing the space bar or after 3 s. Short- and long-delay blocks were randomly intermixed for each participant to ensure that participants would not know which delay to expect on a given block.

During the test phase of each block, 12 pairs were presented at random, one at a time and evenly distributed into intact, related, and unrelated probes. Intact test pairs featured a face–scene pair that had previously been presented during the study phase, such as the young woman paired with the desert in Figure 1a. Related test pairs were recombined face–scene pairs in which the face from one pair was recombined with a scene within the same category. For example, the related probe depicted in the example test phase shown in Figure 1a featured an old man recombined with another kitchen scene from the study phase. Unrelated test pairs were recombined face–scene pairs in which the face from one pair was recombined with a scene from a different category, such as the young woman paired with the kitchen scene in the example test phase shown in Figure 1a. There were 48 of each test probe throughout the experiment, evenly divided between the short- and long-delay blocks. Participants indicated whether a test pair was "intact," "related," or "unrelated"
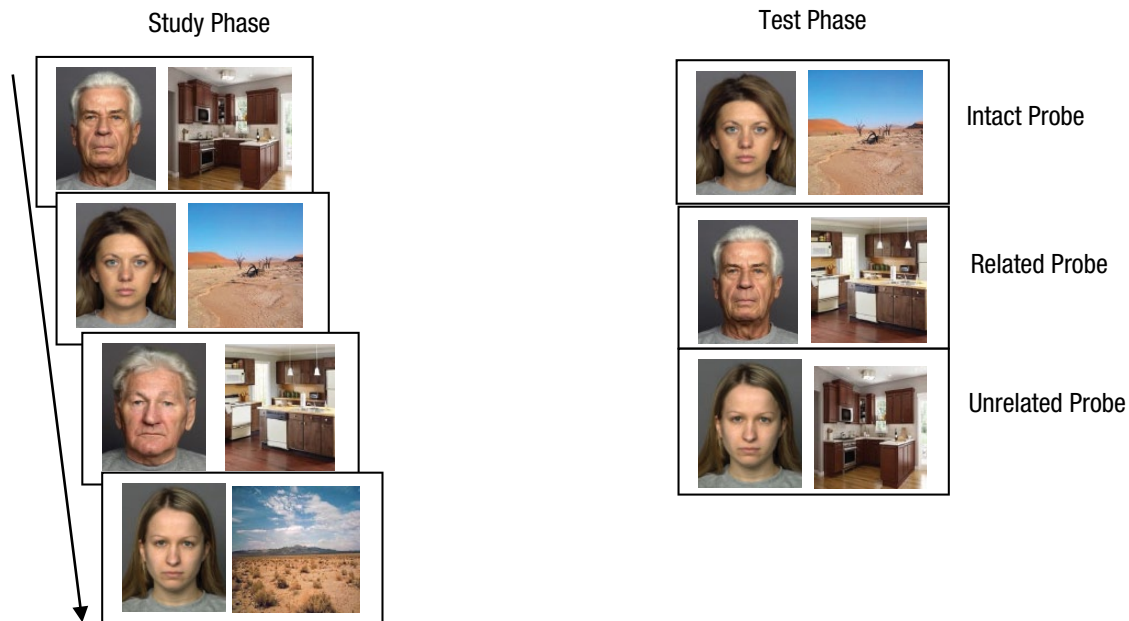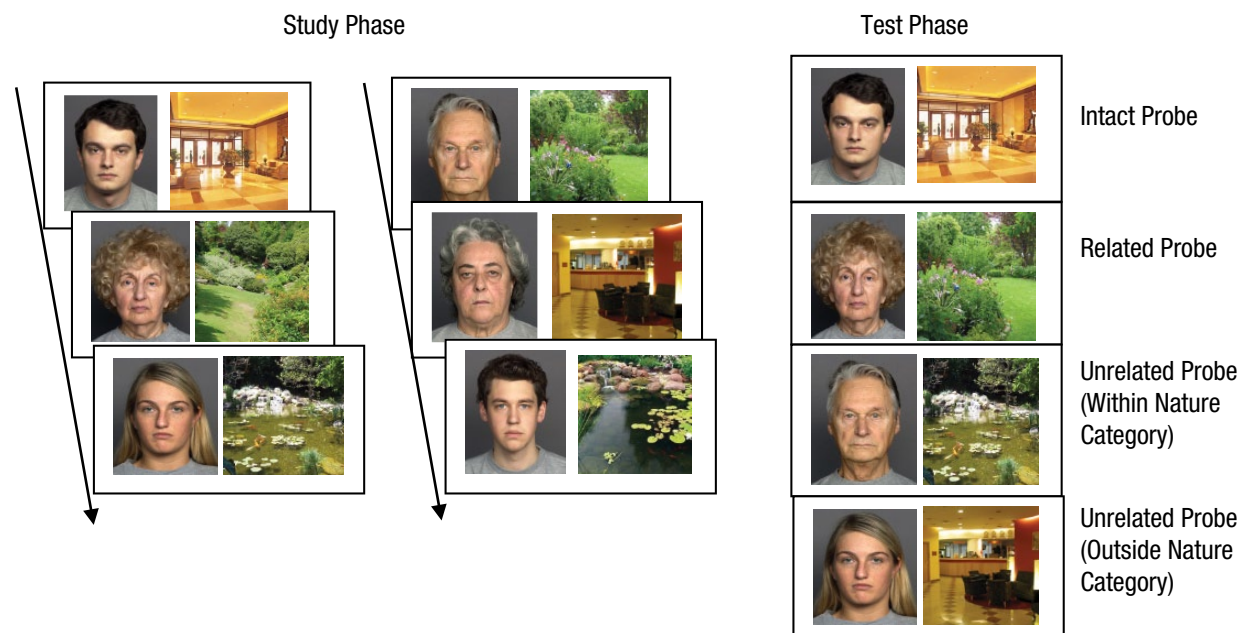
Fig. 1. Illustration of the procedure for Experiment 1 (a) and Experiment 2 (b). After studying a series of face–scene pairs, participants saw 12 test probes. Intact test probes featured a face–scene pair that had previously been presented during the study phase, related test probes featured a face from one pair recombined with a scene within the same category, and unrelated test probes featured a face from one pair recombined with a scene from a different category. In Experiment 2, there were two types of unrelated tests pairs, which varied depending on whether the scene was from the same or a different broader category of indoor or outdoor scenes as in the studied face–scene pair. Participants indicated whether each test pair was "intact," "related," or "unrelated."

by pressing the "a," "g," or "l" key, respectively, which resulted in the next pair appearing. They were untimed in their responses. A 3-s interval separated the blocks. All procedures were approved by the University of Missouri Institutional Review Board.

Results

All data and analysis scripts are available at https://osf .io/xk78c/. Analyses were based on Bayesian posterior estimation, which confers many advantages over null-hypothesis significance testing, such as the ability to quantify evidence for a null effect. Participants' responses to each probe are depicted in Figure 2. As shown, there was evident heterogeneity across participants, which was confirmed with χ^2 tests for participant heterogeneity (Smith & Batchelder, 2008) fitted separately to young- and older adults' data, all χ^2s(145) $\geq$ 627.13, $p < .001$. Therefore, we used hierarchical models for both the logistic regression and MPT analyses reported below.

Logistic regression results. For each probe, we measured age (coded as 0 = young, 1 = older) and delay (0 = short, 1 = long) differences in response accuracy (0 = incorrect, 1 = correct) with hierarchical Bayesian binomial logistic regression models, a more appropriate method of analyzing data on the proportion of correct responses than analysis of variance on aggregated proportions (Dixon, 2008). We also measured age and delay differences in response tendency to each probe (i.e., whether the participant responded "intact," "related," or "unrelated") with categorical logistic regression models. Models were implemented in the *brms* package for R (Bürkner, 2017; R Core Team, 2018). Cauchy priors with location parameters of 0 and scale parameters of 2.5 (for the population-level slopes) and 10 (for the intercepts) were specified (see Gelman, Jakulin, Pittau, & Su, 2008), but we retained the default *t* distribution priors on the random effects. Sampling routine information is available in the Supplemental Material available online (see the section "Hierarchical Bayesian Categorical Logistic Regression Model Diagnostics").

All parameters in logistic regression are on the log-odds scale, so we took the exponent of each estimate to return more easily interpretable odds ratios (*OR*s) and 95% Bayesian credible intervals (*CI*s). The Bayesian *CI* conveys the most probable range of values of the true estimated *OR*; values within the interval have higher credibility than values outside the interval (Kruschke, 2011, 2018). Following the approach to parameter interpretation advocated by Kruschke (2011, 2018), we specified a region of practical equivalence (ROPE) to define a decision threshold to accept or reject a null value. It is important to note that Bayesian

inference is concerned with describing the uncertainty in a parameter rather than hard "black-and-white" decisions to accept or reject values. Nevertheless, we defined a null value (e.g., a null effect of age) as an *OR* of 1, with a ROPE ranging from 0.94 to 1.06. If the resulting 95% CI excluded the ROPE, we rejected the null. If the 95% CI was entirely contained within the ROPE, we accepted the null. Otherwise, we remained agnostic.

First, we assessed whether there were any age differences in accuracy on the task. For intact probes (Fig. 2a), the resulting *OR* comparing older with young adults was 0.63, 95% CI = [0.42, 0.94], indicating that older adults were less accurate. This was likely attributable to older adults' greater tendency (relative to young adults') to respond "related" to intact probes, *OR* = 1.55, 95% CI = [1.07, 2.27]. The evidence was inconclusive as to whether there were age differences in the tendency to respond "unrelated," *OR* = 1.57, 95% CI = [0.84, 3.06], as there were relatively few "unrelated" responses in both age groups. Older adults also responded less accurately to related probes (Fig. 2b), *OR* = 0.36, 95% CI = [0.25, 0.52]. Relative to young adults, older adults were more likely to respond that related probes were "intact," *OR* = 2.86, 95% CI = [1.99, 4.06], or "unrelated," *OR* = 2.59, 95% CI = [1.40, 4.57]. Finally, for unrelated probes (Fig. 2c), the *OR* = 0.61, 95% CI = [0.27, 1.30], partially overlapped with the ROPE and partially excluded it, so we remained agnostic.

Next, regarding delay effects, the evidence for intact probes was inconclusive, *OR* = 0.90, 95% CI = [0.76, 1.06], but note that the modal value lies quite near the ROPE, so there was most probably very little, if any, change in accuracy across delay. However, accuracy for related probes declined modestly with delay, *OR* = 0.77, 95% CI = [0.66, 0.90], with participants being more likely in the long-delay blocks to endorse related probes as "intact," *OR* = 1.39, 95% CI = [1.17, 1.63]. In contrast, for unrelated probes, accuracy slightly improved from short- to long-delay blocks, *OR* = 1.36, 95% CI = [1.11, 1.68], driven by a decreased tendency in the long-delay blocks to endorse unrelated probes as "related," *OR* = 0.74, 95% CI = [0.58, 0.95].

Finally, we tested whether there were any age-by-delay interactions. In all cases, the 95% CI partially overlapped with the ROPE but tended to exclude the ROPE in the tails of the posterior distributions. Therefore, we remained undecided, but a visual inspection of Figure 2 reveals very similar patterns at both the short and long delays.

MPT results. The presence of notable age differences on intact and related probes, but not unrelated probes, suggests that older adults might be deficient relative to young adults at retrieving specific, but not gist, representations of
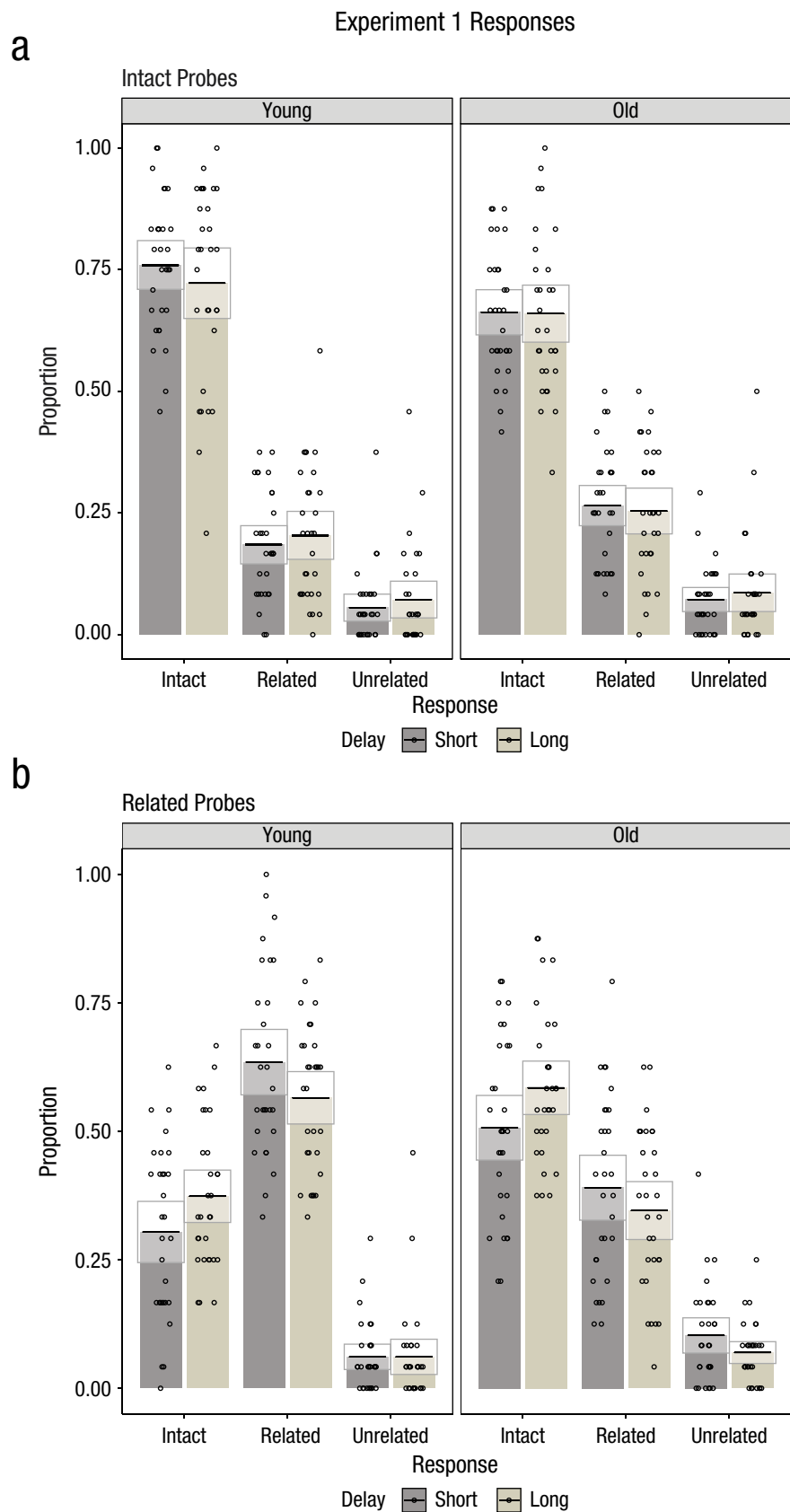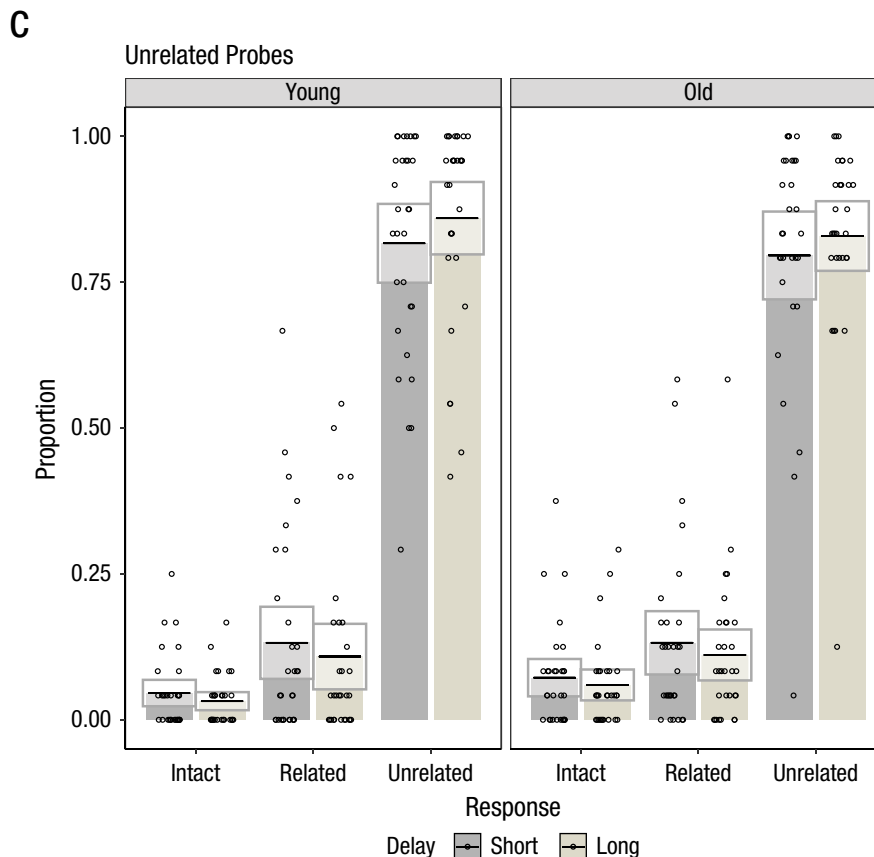
Fig. 2. *(continued on next page)*

C

Unrelated Probes

Fig. 2. Proportion of "intact," "related," and "unrelated" responses to intact probes (a), related probes (b), and unrelated probes (c) in Experiment 1, separately for young and older adults. For each response type, results are shown separately for trials with short and long delays. Lines at the top of the vertical bars denote group means. Boxes around the means indicate 95% confidence intervals. Points denote individual participants' data (jittered for readability).

associations. To test this assumption, we used the MPT model from the SCR paradigm (Stahl & Klauer, 2008), depicted in Figure 3. MPT models attempt to explain how participants arrive at their responses (boxes on the right) to a given memory probe (boxes on the left) by way of different cognitive processes (ovals). The probability of making response j to probe i is expressed as the sum of the probabilities of the branches terminating in that response.

In Figure 3, the two V parameters correspond to the probability that participants retrieve the specific (verbatim) representation of an association given either an intact probe (V_i) or a related probe (V_r). The two G parameters correspond to the conditional probabilities that participants retrieve the gist of an association for intact (G_i) or related (G_r) probes, given that they have not retrieved more specific representations. If participants retrieve the gist, they then guess whether the probe is "intact" (with probability a) or "related" (with complementary probability $1 - a$). If a probe elicits no verbatim or gist information for a participant, then the participant can still guess that the probe is "intact" or

"related" with probability b. Otherwise, with probability $1 - b$, the participant responds that the probe is "unrelated." Note that the unrelated probes contain no verbatim or gist parameters because these probes contain no verbatim or gist representation of originally studied associations, and thus responses to unrelated probes are based entirely on parameter b. Conditional on entering cognitive state b, the participant guesses whether the probe is "intact" (with probability a_b) or "related" (with probability $1 - a_b$). The original SCR paradigm (Stahl & Klauer, 2008) contained a single a parameter for responding "old" whether or not the gist was retrieved. However, this model fitted our data poorly (see the Supplemental Material for details), so we included the additional a_b parameter, which assumed that participants' bias to respond "intact" can change depending on whether or not the gist is retrieved. This modified model had more parameters than degrees of freedom in the data, so we imposed an equality restriction on the gist parameters because the scene paired with a face for both intact and related probes was drawn from the same category. No further parameter
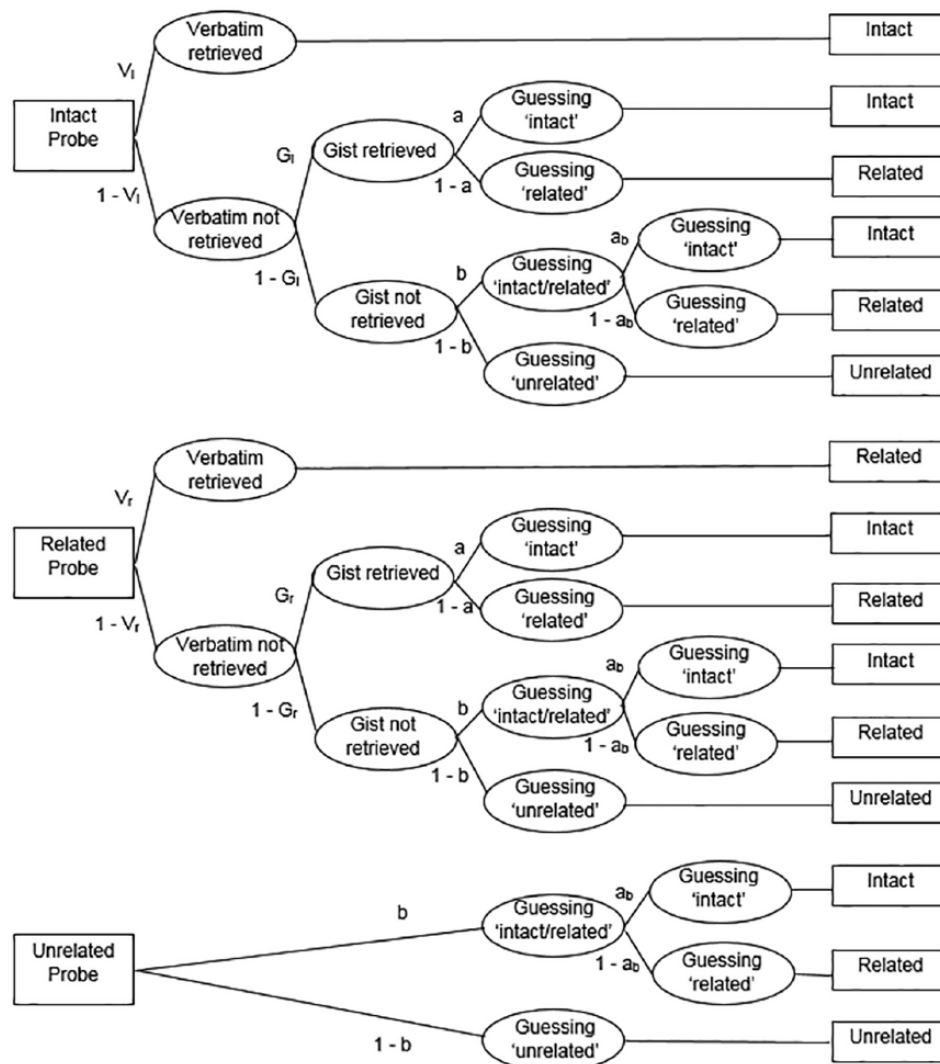
Fig. 3. Multinomial-processing-tree model from the simplified conjoint-recognition paradigm (Stahl & Klauer, 2008). Participants arrive at their responses (boxes on the right) to a given memory probe (boxes on the left) by way of different cognitive processes (ovals). The two V parameters correspond to the probability that participants retrieve the verbatim representation of an association given either an intact probe (V_i) or a related probe (V_r). The two G parameters correspond to the conditional probabilities that participants retrieve the gist of an association for intact (G_i) or related (G_r) probes, given that they have not retrieved more specific representations. If participants retrieve the gist, they then guess whether the probe is "intact" (with probability a) or "related" (with complementary probability $1 - a$). If a probe elicits no verbatim or gist information for a participant, then the participant can still guess that the probe is "intact" or "related" with probability b. Otherwise, with probability $1 - b$, the participant responds that the probe is "unrelated." Conditional on entering cognitive state b, the participant guesses whether the probe is "intact" (with probability a_b) or "related" (with probability $1 - a_b$). The a_b parameter assumes that participants' bias to respond "intact" can change depending on whether or not the gist is retrieved.

restrictions were warranted because each parameter in the model measured a cognitive process of interest and the model was identified (i.e., there were as many parameters as there were degrees of freedom in the data).

We fitted the model separately to both age groups, with hierarchical latent-trait priors (Klauer, 2010) specified using the *TreeBUGS* package for R (Heck, Arnold, & Arnold, 2018; R Core Team, 2018). In both groups,

the 95% CIs for the difference in each parameter between the short- and long-delay blocks overlapped with 0, so we collapsed delay within each group. This approach follows the interpretation of testing for group (or other factor) differences in MPT parameters advocated for by Smith and Batchelder (2010), who argued that if the difference is indistinguishable from 0 (i.e., its CI overlaps with 0), then the parameter does not differ meaningfully across groups. Of course, one could

**Table 1.** Multinomial-Processing-Tree Parameter Estimates

Parameter	Experiment 1		Experiment 2	
	Young adults	Older adults	Young adults	Older adults
V_i	.39 [.18, .56]	.10 [.01, .24]	.59 [.49, .68]	.27 [.14, .39]
V_r	.36 [.21, .48]	.04 [.00, .12]	.27 [.14, .38]	.06 [.00, .13]
$G_i = G_r$	.91 [.86, .95]	.90 [.85, .93]	.63 [.56, .70]	.59 [.52, .64]
a	.63 [.53, .72]	.67 [.59, .75]	.64 [.51, .78]	.62 [.53, .72]
a_b	.34 [.20, .51]	.35 [.25, .45]	.21 [.06, .43]	.29 [.25, .45]
b	.10 [.05, .17]	.15 [.10, .22]	.12 [.06, .17]	.28 [.21, .35]
F			.08 [.03, .13]	.05 [.01, .13]

Note: Estimates are group-level means; values in brackets are 95% Bayesian credible intervals. V_i = verbatim trace is retrieved given an intact probe; V_r = verbatim trace is retrieved given a related probe; G_i = gist trace is retrieved given an intact probe; G_r = gist trace is retrieved given a related probe; a = guessing "intact" on retrieving the gist; a_b = guessing "intact" on entering state b. b = guessing "intact" or "related" when gist is not retrieved; F = a fuzzy level of representation is retrieved.

argue that a ROPE may offer a more detailed test (e.g., Kruschke, 2018), but parameters in MPTs are correlated and interdependent, so a ROPE is likely invalid in this context.

Table 1 provides the parameter estimates for each age group. For each parameter, we subtracted the posterior samples of the older from the young adults to compute a credibility interval of the difference between groups (e.g., Smith & Batchelder, 2010). These differences are presented in Figure 4. Parameters for which the 95% CI of the difference estimate overlaps with 0 do not meaningfully differ between groups. As shown, young adults had higher values of both verbatim-memory parameters (V_i and V_r) but did not differ from the older adults on gist-memory parameters (G_i and G_r) or in response-bias parameters (a, a_b, or b).
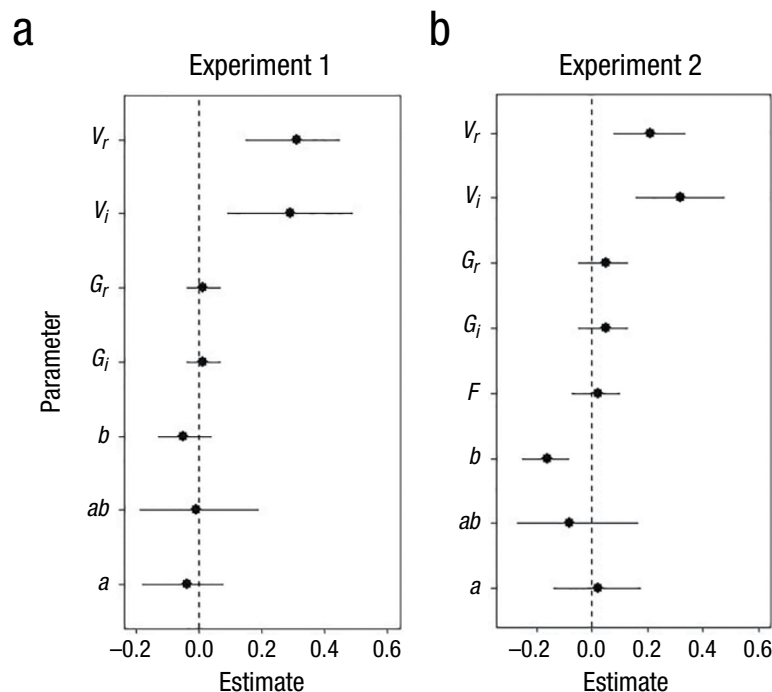
Fig. 4. Forest plots depicting the difference score (young minus older) for each parameter in Experiment 1 (a) and Experiment 2 (b). Points denote posterior mean differences, and error bars show 95% Bayesian credible intervals. The dashed line at 0.0 denotes a point null. a = guessing "intact" when gist is retrieved; a_b = guessing "intact" when gist not retrieved, b = guessing "intact" or "related" when gist is not retrieved; G_i = gist is retrieved given an intact probe; G_r = gist is retrieved given a related probe; F = a fuzzy level of representation is retrieved; V_i = verbatim trace is retrieved given an intact probe; V_r = verbatim trace is retrieved given a related probe.

Because the gist parameters are conditionally reached on failing to access the verbatim trace, the equivalent gist parameters indicate that both young and older adults are equally likely to retrieve the gist of an association when they fail to access the verbatim trace. However, to derive a composite score of how often participants retrieve some degree of representation for an original association, we must combine information about how often participants access either the verbatim or the gist traces. This can be achieved with the following equations— for intact probes: $V_i + (1 - V_i) \times G_i$; for related probes: $V_r + (1 - V_r) \times G_r$. Plugging in parameter estimates from Table 1, we derive composite retrieval scores for the intact probes of .95, 95% CI = [.89, .98], for the young adults and .91, 95% CI = [.85, .95], for the older adults. For the related probes, composite retrieval scores are .94, 95% CI = [.89, .97], for the young adults and .90, 95% CI = [.85, .94], for the older adults. That is, older adults can retrieve at least some degree of representation of associations as well as the young adults, but they are impaired in their ability to retrieve the most specific levels of representation.

Signal detection theory (SDT) analyses. We also modeled our data using the equal-variance SDT model, given an ongoing debate about whether recognition proceeds on a discrete or continuous basis (Batchelder & Alexander, 2013; Dube & Rotello, 2012; Pazzaglia, Dube, & Rotello, 2013). This debate has been somewhat extended to associative recognition (see Rotello, 2017) and to conjoint-recognition tasks (Brainerd, Gomes, & Moran, 2014). The SDT results are presented in detail in the Supplemental Material and generally converge with our MPT results in both experiments (see Table S1). One notable exception is that the SDT results suggested that older adults in Experiment 1 had a more liberal bias (i.e., were more likely to respond "intact") when comparing intact with related probes, whereas the MPTs showed no age-related differences in response-bias parameters. This is likely attributable to the fact that in the MPT model, there are several processes that contribute to responding "intact," including retrieving verbatim memory, guessing "intact" when only the gist is retrieved, and guessing "intact" when neither the verbatim nor the gist is retrieved. In contrast, in our SDT models, the only parameters were memory strength and response bias.

Discussion

These results support our hypothesis that age-related deficits in associative memory are restricted to the retrieval of specific, but not gist, representations. Older adults made more errors than young adults on intact and related, but not unrelated, probes, and the fit of an MPT model demonstrated that older adults had lower estimates of verbatim, but not gist, memory.

There was no effect of delay on verbatim or gist memory for associations. It is possible that the retention intervals used in the long-delay blocks were not long enough to show any continued loss of verbatim memory or potential loss of gist memory, so we extended the retention interval in Experiment 2.

We categorized face–scene pairs to strengthen schematic support for a given association by offsetting alternative schemas, which might help with retrieving the gist. However, this created an imbalance in the number of specific associations participants needed to maintain per block (12; e.g., Old Man 1–Park 1 and Old Man 2–Park 2) compared with the number of gist associations to be maintained (6; e.g., old man–park). Therefore, in Experiment 2, we removed categorized pairings. We also introduced a novel manipulation to measure different levels of specificity for associations.

Experiment 2

Method

Participants. We recruited 40 young adult participants (72.5% female; age: range = 18–24 years, $M = 19.68$, $SD = 1.86$) and 40 older adult participants (75% female; age: range = 64–80 years, $M = 73.05$, $SD = 3.99$). Young adults were compensated with research credits for psychology courses, and older adults were compensated with $15 for their time and travel expenses. Data from an additional 3 young and 4 older adults are not reported in these analyses because they performed well below chance level (i.e., < 33% accuracy across the different probe conditions). Our sample size was based on the same approach as used in Experiment 1, but because the MPT parameters were estimated from fewer trials per participant, we increased our sample size to 40 participants to more precisely estimate the parameters. The parameter estimates were robust across different Bayesian-prior specifications (see https://osf.io/fpcmu/).

Stimuli and procedure. Stimuli were identical to those used in Experiment 1, but we created only 84 face–scene pairs; 72 pairs were presented during the experimental blocks, and 12 were presented in a practice block. Although the procedure was mostly similar to those used in Experiment 1 (see Fig. 1b), there were a few noteworthy changes. First, we reduced the total number of blocks from 12 to 6 to keep the experimental session to under an hour. Three short-delay and three long-delay blocks were randomly intermixed for each participant, and the verbal semantic-memory filler task was extended to occupy 301 s (~5 min) between the study and test phases

in the long-delay blocks. This reduced the number of test probes to 24 of each type (i.e., intact, related, and unrelated), evenly divided between short and long delay.

Second, we equated the number of specific and gist associations to be maintained in a single block by removing categorized pairings. Although two exemplars of a given scene type (e.g., two parks; see Fig. 1b) both appeared in the same block, we ensured that each exemplar was paired with a different age- or gender-categorized face (e.g., Old Woman 1–Park 1, Old Man 1–Park 2). However, related test pairs were still constructed by recombining a face with a similar scene (e.g., Old Woman 1–Park 2; see the test phase in Fig. 1b).

Third, we organized the scenes in each block such that half of the scenes belonged to a broader nature category (e.g., two parks, two forests, two streams) and half to a broader indoor category (e.g., two corridors, two lobbies, two dens). This manipulation was designed to test the possibility for retrieval of different levels of specificity for associations; that is, whether an individual can retrieve an original association (e.g., "This old man was with this park"), a less-specific representation that still retains information about the type of scene paired with a given face (e.g., "This old man was with some park"), or a fuzzier representation retaining information about the broader scene category paired with a given face (e.g., "This old man was outside somewhere"). To test this, we constructed the unrelated test pairs such that half of them featured a scene switch from within the same broader category (e.g., from Old Man 1–Park 2 to Old Man 1–Stream 1, both nature scenes; unrelated-within category), and half featured a scene switch from outside the broader category (e.g., from Young Woman 1–Stream 1 to Young Woman 1–Lobby 2; unrelated-opposite category). The procedure was otherwise identical to that in Experiment 1.
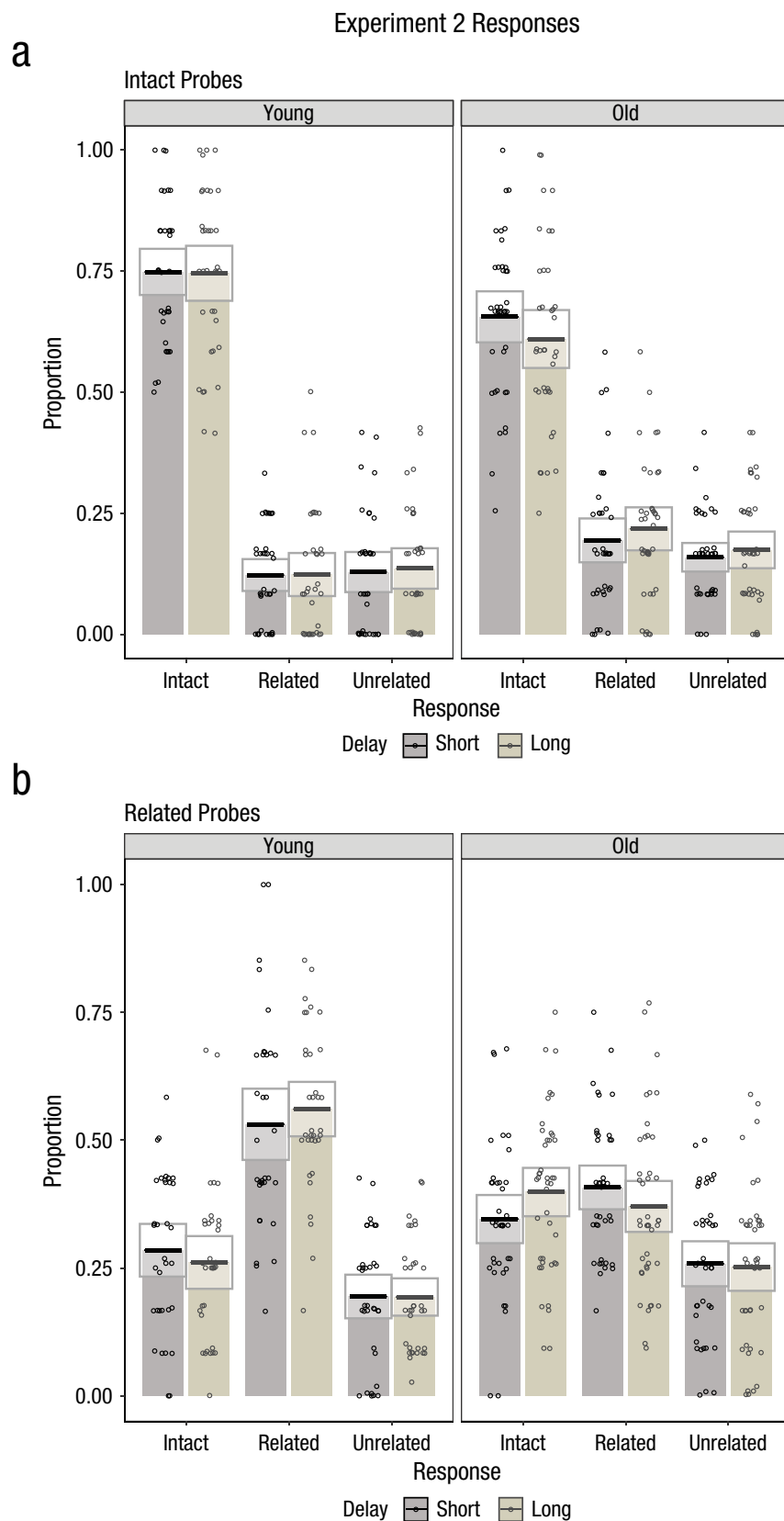
Results

Logistic regression results. The proportion of responses to each probe is depicted in Figure 5. There was substantial heterogeneity in responses in both age groups, all χ^2s(312) $\geq$ 636.21, $p < .001$. The age effects were mostly consistent with Experiment 1. Specifically, older adults were less accurate than young adults on intact probes, $OR = 0.47$, 95% CI = [0.32, 0.69]. However, whereas in Experiment 1 this was driven by older adults' tendency to endorse intact probes as "related," in Experiment 2, they were not only more likely than young adults to respond "related," $OR = 2.48$, 95% CI = [1.57, 3.93], but also more likely to respond "unrelated" to intact probes, $OR = 1.93$, 95% CI = [1.25, 3.00]. As in Experiment 1, older adults were less accurate than young adults on related probes, $OR = 0.48$, 95% CI = [0.36, 0.64], and were more likely

than young adults to endorse related probes as either "intact," $OR = 2.20$, 95% CI = [1.62, 2.97], or "unrelated," $OR = 1.84$, 95% CI = [1.36, 2.44]. Unlike in Experiment 1, there was clear evidence in Experiment 2 that older adults were less accurate than young adults on unrelated probes, $OR = 0.45$, 95% CI = [0.28, 0.73], and were more likely than young adults to endorse unrelated probes as "intact," $OR = 2.83$, 95% CI = [1.58, 5.10], or "related," $OR = 1.93$, 95% CI = [1.12, 3.42].
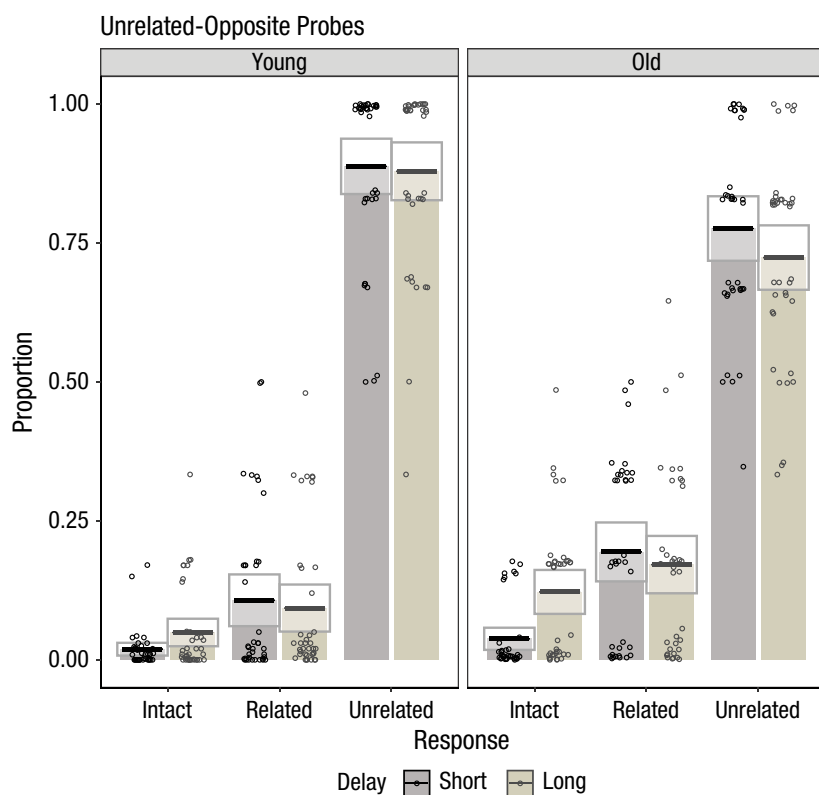
For the unrelated probes, we also tested whether performance differed between the two types of unrelated probes (coded as 0 = unrelated-within, 1 = unrelated-opposite). The age effect held for both types. However, across both age groups, accuracy was lower for the unrelated-within probes, $OR = 0.61$, 95% CI = [0.49, 0.76]. This could reflect that some participants retrieved a fuzzier level of representation for these probes, which maintained some gistlike similarity to the original associations. We tested this possibility with the MPT analyses described in the next section. Finally, for all probes, we remained agnostic about any effects of delay or age-by-delay interactions because in all cases, the 95% CI for the OR mostly overlapped with the ROPE, but the tails of the posterior distributions tended to marginally exclude the ROPE. A visual inspection of Figure 5 reveals very little, if any, changes in performance across delay conditions.

MPT results. We expanded the MPT model shown in Figure 3 to include two separate trees for unrelated probes. For the unrelated-opposite probes, the tree is identical to the unrelated tree in Figure 3. For the unrelated-within probes, an additional parameter (F) corresponded to retrieving a fuzzier level of representation; specifically, with probability F, a participant thought the probe was similar to an original association (e.g., "This old man was somewhere in nature") and then guessed whether the probe was "intact" (with probability a) or "related" (with probability $1 - a$). With probability $1 - F$, the participant did not retrieve a fuzzy level of representation, and the tree resolved to the unrelated tree shown in Figure 3. We retained the equality restraint on the gist parameters (G_i and G_r) to make the model with the new F parameter overidentified. Tests of model fit with and without the equality restraint are reported in the Supplemental Material.

We collapsed the delay conditions within each age group because the 95% CIs of the difference in each parameter between the short- and long-delay blocks overlapped with 0. Table 1 provides a summary of parameter estimates, and Figure 4b depicts the difference score obtained by subtracting the posterior samples of the older from the young adults. These results are in line with those of Experiment 1. There were age differences on the verbatim-memory parameters (V_i and
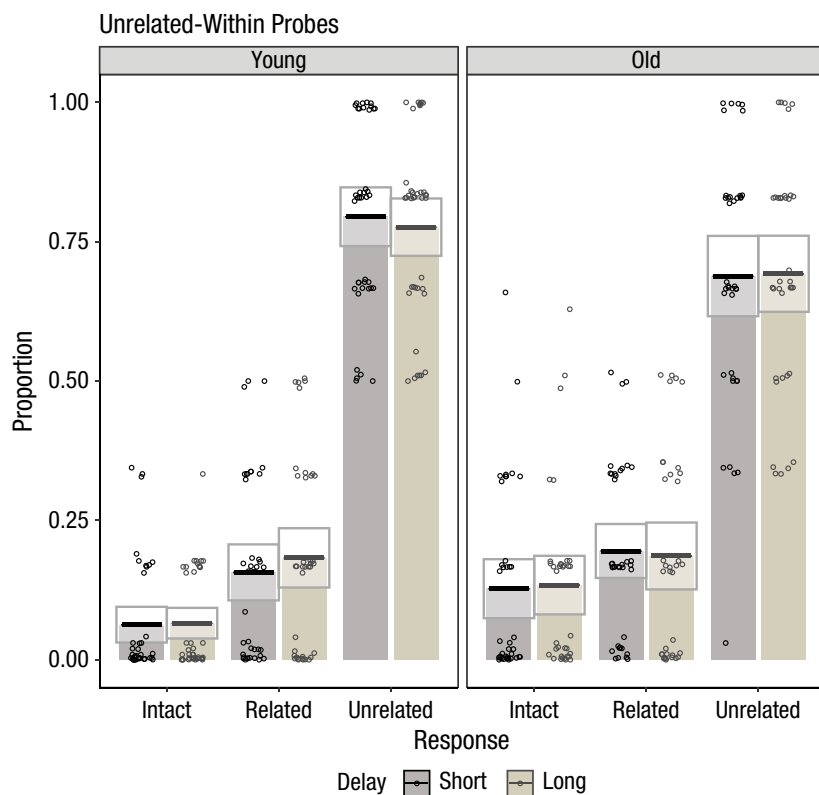
Experiment 2 Responses

a

Intact Probes

b

Related Probes

Fig. 5. *(continued on next page)*

c

Unrelated-Opposite Probes

d

Unrelated-Within Probes

Fig. 5. Proportion of "intact," "related," and "unrelated" responses to intact probes (a), related probes (b), unrelated-opposite probes (c), and unrelated-within probes (d) in Experiment 2, separately for young and older adults. For each response type, results are shown separately for trials with short and long delays. Lines at the top of the vertical bars denote group means. Boxes around the means indicates 95% confidence intervals. Points denote individual participants' data (jittered for readability). Note that splitting the unrelated probes into two categories resulted in fewer trials per category, so the distribution of scores appears more clustered than for intact and related probes.

V_r) but not the gist-memory parameters (G_i and G_r). Older adults had a higher bias to respond "intact" or "related" to unrelated probes (parameter b), which is likely attributable to the fact that they made more errors on unrelated probes than the young adults did. Notably, the model fitted the data well, with the addition of a fuzzy-retrieval parameter (F) in both age groups, which supports the notion that some participants may have retrieved a less specific level of representation for some of the unrelated-within probes. This parameter did not differ between groups.

We used the approach outlined in Experiment 1 to derive composite retrieval scores for the intact and related probes. Plugging in values from Table 1, we derive composite retrieval scores for the intact probes of .85, 95% CI = [.78, .90], for the young adults and .70, 95% CI = [.59, .78], for the older adults. For the related probes, composite retrieval scores were .73, 95% CI = [.62, .81], for the young adults and .61, 95% CI = [.52, .69], for the older adults. There are two important aspects of these scores. First, older adults generally had lower overall retrieval scores than the young adults. Second, composite retrieval scores in Experiment 2 were generally lower than in Experiment 1. This suggests that when schematic support at encoding is reduced, participants retrieve less information about originally encoded associations, and this is especially true of older adults. Importantly, older adults exhibit the greatest deficit in retrieval of highly specific information (verbatim memory), show a less pronounced deficit in verbatim-plus-gist retrieval (composite memory score), and show relatively equivalent retrieval of fuzzy levels of specificity as the young adults (parameter F in the MPT model). In other words, the age-related deficits in associative memory observed in the present study scale with the amount of specificity needed to be retrieved, consistent with the specificity principle of memory.

Discussion

Results from Experiment 2 were mostly consistent with and extended those of Experiment 1. Older adults exhibited the largest deficits in retrieval of associations at highly specific levels (e.g., "The old man was in *this*

park"). They also demonstrated slightly lower composite retrieval (verbatim-plus-gist memory). Proportionally, they retrieved the gist of associations (e.g., "The old man was in *some* park"), on failing to retrieve the specific representation, to similar extents as young adults, but the pronounced loss of verbatim memory resulted in lower overall composite retrieval scores. However, they retrieved fuzzier levels of representation (e.g., "The old man was outside somewhere") to the same extent as young adults. These findings point to a specificity gradient in associative episodic memory, whereby more specific information is harder to retrieve, especially with age, whereas representations are relatively more preserved at lower levels of specificity. This is consistent with Craik's (2002) hierarchical representation model, but it extends that model by offering firmer evidence for multiple levels of representation than has been shown in any previous study.

An important distinction between the two experiments was the schematic support manipulation. When pairings were categorized (Experiment 1), composite retrieval scores were higher in both age groups than when this manipulation was absent (Experiment 2). The removal of categorized pairings made it possible that even the gist of an association might become susceptible to interference or forgetting. This appeared to occur to some extent in Experiment 2 for both young and older adults, a finding compatible with the specificity principle of memory (Surprenant & Neath, 2009).

As with Experiment 1, we found no delay effect on verbatim or gist memory. This supports the notion that difficulties with retrieving specific information occur on multiple time scales, which is central to the universality of principles of memory (Surprenant & Neath, 2009). However, future research is warranted to assess whether specific or gist memory for associations declines with longer delays than those used in the present experiments.

General Discussion

Guided by Surprenant and Neath's (2009) specificity principle of memory, we have shown that older adults, relative to young adults, are deficient at retrieving associative episodic memories at highly specific levels of

representation but can retrieve associations at lower levels of specificity (e.g., gist) to relatively similar extents as young adults. To date, most of the evidence suggesting that older adults rely on a gist-based processing strategy has come from studies employing categorized lists of material in which a semantic gist may be activated because all of the items are related (see Devitt & Schacter, 2016; Tun, Wingfield, Rosen, & Blanchard, 1998). Our findings extend these claims by showing that older adults can remember the gist of associations, which comprise the core of episodic memory (Tulving, 1983), providing perhaps the most direct evidence to date that older adults retrieve fuzzier episodic representations. Moreover, our findings of multiple levels of specificity for associations suggest that episodic memory representations might exist on a continuum (cf. Craik, 2002).

Although we suggest that older adults struggle to retrieve specific information, it is also possible that they may fail to encode specific information (e.g., Naveh-Benjamin, 2000). However, Luo and Craik (2009) found that age-related deficits in memory for specific information are more likely attributable to retrieval than to encoding failures, but it is imperative for future research to continue to disentangle the effects of encoding versus retrieval specificity. In our view, it is likely a dynamic relationship between encoding and retrieval that influences older adults' memory for specific information, a view shared by Surprenant and Neath (2009). Further, as with any aging study employing a cross-sectional design, we cannot state that aging causes a loss of specific memory for associations, because other factors may be at fault.

These results are consistent from both an MPT and a signal detection framework. This is important, given a growing debate as to whether recognition proceeds on a discrete or continuous basis (e.g., Pazzaglia et al., 2013). One important area in which this debate has not been extended much is the simplified conjoint-recognition procedure (Stahl & Klauer, 2008), the type of paradigm used here. More work is also warranted in extending this debate to associative recognition (e.g., Rotello, 2017), especially in conjunction with the measurement of verbatim and gist memory for associations.

We used face–scene pairs because they simulate remembering where people were encountered (e.g., Gruppuso et al., 2007). Memory for where events occurred is one of the hallmarks of episodic memory (Tulving, 1983). The extent to which these findings would generalize to other types of associations in memory cannot be directly answered from the present study but would seem reasonable considering the similar decline they show with age (cf. Old & Naveh-Benjamin, 2008). For some associations, however, such as remembering who committed a particular action, retrieval of the gist may not be sufficient, given potential legal implications for eyewitness testimony. Nevertheless, our findings that older adults can retrieve the gist of where they previously encountered people indicates that, at least for some associations in memory, age-related deficits are limited to retrieval of the specific representations.

ORCID iD

Nathaniel R. Greene (iD) https://orcid.org/0000-0001-6758-6761

Supplemental Material

Additional supporting information can be found at http:// journals.sagepub.com/doi/suppl/10.1177/0956797620901760

References

Batchelder, W. H., & Alexander, G. E. (2013). Discrete-state models: Comment on Pazzaglia, Dube, and Rotello (2013). *Psychological Bulletin*, *139*, 1204–1212.

Boywitt, C. D., Kuhlmann, B. G., & Meiser, T. (2012). The role of source memory in older adults' recollective experience. *Psychology and Aging*, *27*, 484–497.

Brainerd, C. J., Gomes, C. F. A., & Moran, R. (2014). The two recollections. *Psychological Review*, *121*, 563–599.

Brainerd, C. J., & Reyna, V. F. (2015). Fuzzy-trace theory and lifespan cognitive development. *Developmental Review*, *38*, 89–121.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). doi:10.18637/jss.v080.i01

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging*, *20*, 718–721.

Castel, A. D., McGillivray, S., & Worden, K. M. (2013). Back to the future: Past and future era-based schematic support and associative memory for prices in younger and older adults. *Psychology and Aging*, *28*, 996–1003.

Chen, T., & Naveh-Benjamin, M. (2012). Assessing the associative deficit of older adults in long-term and short-term/ working memory. *Psychology and Aging*, *27*, 666–682.

Craik, F. I. M. (2002). Levels of processing: Past, present . . . and future? *Memory*, *10*, 305–318.

Devitt, A. L., & Schacter, D. L. (2016). False memories with age: Neural and cognitive underpinnings. *Neuropsychologia*, *91*, 346–359.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, *59*, 447–456.

Dube, C., & Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 130–151.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavioral Research Methods*, *42*, 351–362.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Minimental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*, 1360–1383.

Gruppuso, V., Lindsay, D. S., & Masson, M. E. J. (2007). I'd know that face anywhere! *Psychonomic Bulletin & Review*, *14*, 1085–1089.

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavioral Research Methods*, *50*, 264–284.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*, 70–98.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*, 1551–1556.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*, 299–312.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*, 270–280.

Luo, L., & Craik, F. I. M. (2009). Age differences in recollection: Specificity effects at retrieval. *Journal of Memory and Language*, *60*, 421–436.

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.

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.

Pazzaglia, A. M., Dube, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, *139*, 1173–1203.

R Core Team. (2018). R: A language and environment for statistical computing (Version 3.4.4) [Computer software]. Retrieved from http://www.R-project.org/

Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making*, *4*, 249–262.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75.

Roediger, H. L., III. (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology*, *59*, 225–254.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.

Rotello, C. M. (2017). Signal detection theories of recognition memory. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed.). San Diego, CA: Academic Press.

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*, 301–308.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime 2.0 reference guide manual*. Pittsburgh, PA: Psychology Software Tools.

Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review*, *15*, 713–731.

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*, 167–183.

Stahl, C., & Klauer, K. C. (2008). A simplified conjoint recognition paradigm for the measurement of gist and verbatim memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 570–586.

Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. New York, NY: Psychology Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford, England: Clarendon Press.

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