



Different patterns of recollection for matched real-world and laboratory-based episodes in younger and older adults

Nicholas B. Diamond^{a,b,d,*}, Hervé Abdi^e, Brian Levine^{a,b,c,**}

^a Rotman Research Institute, Baycrest Health Sciences, Toronto, Ontario M6A 2E1, Canada

^b University of Toronto, Department of Psychology, Toronto, Ontario M5S 3G3, Canada

^c University of Toronto, Department of Medicine (Neurology), Toronto, Ontario M5S 3H2, Canada

^d University of Pennsylvania, Department of Psychology, Philadelphia, PA 19104, USA

^e School of Behavioral and Brain Sciences, The University of Texas at Dallas, Dallas, TX 75080, USA

ARTICLE INFO

Keywords:

Episodic memory

Autobiographical memory

Aging

ABSTRACT

To bridge the gap between naturalistic and laboratory assessments of episodic memory, we designed time- and content-matched real-world and virtualized versions of the same tour event. In younger and older adults, we investigated objective and subjective aspects of recollection for event features using a verbal true/false test common to both event conditions. Using a data-driven multivariate analysis blind to the age groups and event conditions, we found that discriminating altered from true details accounted for the largest amount of variance in objective retrieval patterns. There was an advantage for real-world over laboratory encoding on this dimension for both age groups. Similarly, real-world encoding elicited higher scores on a dimension defined by subjective recollection. However, real-world (but not laboratory) encoding decoupled objective and subjective memory in older adults, who reported similar rates of subjective recollection as younger adults despite exhibiting significantly poorer discrimination accuracy. These results demonstrate robust and specific ways in which the accuracy and subjective quality of memory differ for matched naturalistic and laboratory episodes. Furthermore, these results suggest that naturalistic and laboratory encoding conditions produce qualitatively different patterns of episodic memory decline in older age.

1. Introduction

Events are the “prototypical unit[s]” of episodic memory (Tulving, 1983, p. 223), but they are operationalized in dramatically different ways across studies, ranging from lists of discrete stimuli in typical laboratory studies to personal real-world experiences in autobiographical memory studies. The relative merits of laboratory versus naturalistic memory assessments have been debated for decades (e.g. Banaji & Crowder, 1989; Conway, 1991; Koriat & Goldsmith, 1996; Neisser, 1978; Tulving, 1991), with the former emphasizing experimental control and the latter emphasizing ecological validity (for a review, see Kvavilashvili & Ellis, 2004). There has been an increase in naturalistic memory studies over this time (see Fig. S1), but direct comparison is required to determine whether there are theoretically important differences in the memories produced by naturalistic versus laboratory events.

Relative to laboratory stimuli, autobiographical events tend to be

more remote (and therefore span a larger search space; Chen, Gilmore, Nelson, & McDermott, 2017) as well as more emotional, personally significant, and self-referential (Cabeza & St. Jacques, 2007). Though these variables can be manipulated within conventional laboratory memory paradigms, certain features of autobiographical and laboratory events remain inherently different – for instance: embodiment, the contribution of multisensory and idiothetic cues, and the spatial scale.

Consistent with these suggested differences, brain network patterns during retrieval of autobiographical versus laboratory-based events often diverge, with the former eliciting greater activity or discriminability in the hippocampus, posterior medial temporal lobe and cortical midline regions of the default mode network (Cabeza et al., 2004; Chen et al., 2017; Chow, Westphal, & Rissman, 2018; Monge, Wing, Stokes, & Cabeza, 2018; Summerfield, Hassabis, & Maguire, 2009; for earlier meta-analyses, see Gilboa, 2004; McDermott, Szpunar, & Christ, 2009). These findings echo differences in hippocampal coding of real-world versus virtual experiences in rodents and monkeys (Aghajanian et al.,

* Correspondence to: N. B. Diamond, Department of Psychology, Philadelphia, PA 19104, USA.

** Correspondence to: B. Levine, Rotman Research Institute, Baycrest Health Sciences, Toronto, Ontario M6A 2E1, Canada.

E-mail addresses: diamondnb@sas.upenn.edu (N.B. Diamond), blevine@research.baycrest.org (B. Levine).

2015; Thome et al., 2017), together suggesting that the processes supporting memory for laboratory versus naturalistic events may be fundamentally different (Roediger & McDermott, 2013).

On the other hand, when laboratory methods emphasize contextual recall or subjective re-experiencing, the associated pattern of brain activation is similar to that engaged by autobiographical memory (Kim, 2015; Rissman, Chow, Reggente, & Wagner, 2016; Rugg & Vilberg, 2012). This raises the possibility that observed neural-level differences in memory for laboratory versus naturalistic events reflect incidental task features (e.g. retrieval cues and instructions, retention interval, and event content itself) rather than crucial component processes. Yet, holding constant incidental task features, it is still unknown whether and how memories differ across naturalistic and laboratory encoding contexts in the first place. Such differences could be observed in the objective and subjective criteria of episodic memories (respectively, accurate detail retrieval and subjective re-experiencing).

Evidence from the spatial memory literature suggests that first-person real-world encoding conditions enhance memory relative to more passive or cue-impooverished conditions. For instance, real-world versus desktop-based (video or virtual) navigation produce qualitatively different spatial memory representations (Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006), and active self-motion produces better spatial relational memory than passive displacement (Holmes, Newcombe, & Shipley, 2018). It is unclear if similar manipulations of environmental context and physical engagement shape the accuracy and vividness of memory for events more generally, despite established evidence for the scaffolding effects of space on episodic memory (Robin, 2018).

If naturalistic and laboratory episodic memories differ in crucial ways, they may also exhibit different patterns of decline across groups. Older age, in particular, is associated with episodic memory changes due to degradation in the extended hippocampal circuit (Leal & Yassa, 2015; Nilsson et al., 1997). Relative to younger adults, older adults recall fewer perceptual and contextual details from past events and typically report less subjective re-experiencing (Hashtroudi, Johnson, & Chrosniak, 1990; Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002; Old & Naveh-Benjamin, 2008; Parkin & Walter, 1992). It is, however, not clear how aging interacts with memory for naturalistic versus laboratory encoding contexts.

By some accounts, features of real-world encoding conditions such as spatial contextual distinctiveness (Sharps & Gollin, 1987; Uttl & Graf, 1993), environmental support (Craik, 1986), and active participation (Hashtroudi, Parker, Luis, & Reisen, 1989) should reduce age-related episodic memory deficits. On the other hand, increased irrelevant and distracting information in real-world settings should selectively impair older adults (Hasher & Zacks, 1988). Furthermore, when volitional control affords active and adaptive encoding strategies, older adults may be less likely than younger adults to adopt these strategies (Brandstatt & Voss, 2014), and less likely to bind and exploit richer pools of available contextual information to support recollection (Chalfonte & Johnson, 1996; Hay & Jacoby, 1999). In still other cases, increasing stimulus distinctiveness or enactive encoding improves younger and older adults' memories similarly (Park, Cherry, Smith, & Lafronza, 1990; Rosa & Gutchess, 2011; Schacter, Israel, & Racine, 1999).

In the present study, we had three goals: first, to bring methodological rigour to naturalistic memory assessment by objectively measuring episodic memory for a rich and dynamic yet experimentally controlled event, building on other recent controlled real-world encoding paradigms (e.g. Armson, Abdi, & Levine, 2017; Dede, Frascino, Wixted, & Squire, 2016; Diamond, Romero, Jeyakumar, & Levine, 2018; Jeunehomme & D'Argembeau, 2018; St. Jacques & Schacter, 2013). Second, to experimentally compare episodic memory retrieval, including its objective and subjective components, for time- and content-matched naturalistic and laboratory-style episodes using an identical test. Third, to investigate how aging, associated with episodic memory

decline, interacts with this encoding manipulation. We designed our encoding conditions, memory assessment, and analytical approach to identify which aspects of episodic retrieval reliably differed across encoding conditions and age groups. Our approach was exploratory with regard to the precise nature of these differences.

We created real-world and virtualized (slideshow) versions of the same event: a tour of the arts and exhibits on the first floor of Baycrest Health Sciences, which is a visually rich and interesting environment. Both encoding conditions were yoked to the same museum-style audio guide, which cued participants' attention to specific target items, controlling viewing time and order. After a 48 h delay, all participants completed the same true/false memory test with accompanying subjective ratings probing objective and subjective recollection for specific details from the tour. We used a multivariate data reduction technique (multiple correspondence analysis; MCA), that was blind to the encoding conditions and age groups, to identify (1) the patterns of test responses that characterize the main dimensions of variability in the data, and then (2) whether the groups differ along these dimensions. This approach provided a statistically powerful and data-driven method to investigate whether and how age and encoding context (naturalistic versus laboratory) affect episodic memory accuracy and phenomenology.

2. Methods

2.1. Participants

Participants were 84 adults, 43 younger ($M_{\text{age}} = 23.38$ years, $SD = 3.75$; $M_{\text{education}} = 15.50$ years, $SD = 1.31$; 27 women) and 41 older ($M_{\text{age}} = 69.95$, $SD = 3.81$; $M_{\text{education}} = 16.35$ years, $SD = 3.01$; 28 women). They were recruited via the Rotman Research Institute participant database at Baycrest Health Sciences, and from advertisements in the Toronto community. Four participants were excluded; three (one younger and two older) for abnormally poor neuropsychological test performance, and one older adult for failing to adhere to study guidelines by revisiting part of the tour (see below) between encoding and test sessions. Participants were screened for history of neurological or psychiatric illness, active significant medical illness, or substance abuse. Given that the encoding conditions in the present study involved either seated computer use or physical ambulation, all participants were screened for experience using computers and comfort standing and walking for 30 min. Participants were fluent English speakers, had normal or corrected-to-normal vision and hearing (by self-report), were not colour-blind, and gave informed consent in accordance with institutional guidelines.

2.2. Materials

2.2.1. The event conditions

The naturalistic tour (NAT) condition involved an audio-guided real-world walking tour of the first floor of Baycrest hospital (see Fig. 1A). We designate this event 'Baycrest Tour 1.0', as distinguished from different tours used in subsequent studies. The route formed a loop through several different sections of the building. Although participants were instructed to approach different target items, the tour was generally unidirectional. Participants were instructed to examine target items (e.g. paintings, portraits, and exhibits) and to complete different tasks (e.g. locate a particular individual in a frame of portraits, or to locate a particular item in the gift shop). In the middle of the tour, participants had an interaction with a research confederate, during which the confederate asked a series of scripted questions.

The laboratory tour (LAB) consisted of a slideshow capturing the content of the naturalistic tour in a series of static images (see Fig. 1B). It was controlled by the same audio guide as the NAT tour. Participants viewed a slideshow of colour photographs presented using Microsoft PowerPoint. The photographs were taken using a Fujifilm X20 digital



Fig. 1. A. In the real-world condition (NAT), participants underwent a walking-tour of the first floor of Baycrest Hospital. The tour formed a loop. Participants listened to a museum-style audio guide through headphones. The guide instructed participants to examine certain artwork and items (two examples are shown), controlling viewing time and order. B. In the laboratory condition (LAB), participants were exposed to a virtualized version of the real-world tour captured in discrete photographs. LAB participants also listened to the audio guide from NAT.

camera. The LAB slideshow consisted of 106 photographs that can be classified as scene shots, item shots, or buffer shots. Scene shots depicted hallways or rooms as they would be seen by participants at the corresponding point in NAT. Item shots depicted target items in full screen and were linked to their appropriate audio tracks. Buffer shots depicted non-target items in full screen; they were included in an effort to equate total encoding duration and intervening non-target event content across conditions. The photographs appeared in sequence

according to the order in which the content was encountered in the NAT condition, and the number of scene and buffer shots between each target item was scaled to approximate the distances between target items. For the confederate interaction, participants in the LAB condition viewed a picture of the confederate and read the same scripted questions, then answered silently. Participants advanced through the slideshow by pressing the spacebar.

The audio guide was recorded and edited by N.B.D using Audacity

(<http://audacity.sourceforge.net/>). It is publicly available on the Open Science Framework (<https://osf.io/pmt7d/>). It is narrated by four different speakers (two female and two male). Participants listened to the guide using Creative HQ-1600 over-ear headphones from a portable MP3 player (NAT) or a testing laptop (LAB). Each speaker narrated two different sections of the tour, and the narrator order was different in sections 1-4 vs. 5-8, so that the narrator order was decoupled from the sequence structure of the tour. Several measures were taken to increase the distinctiveness of the four narrators. Participants in both conditions heard narrator biographies before the tour began, and twice more throughout the tour. They also received a sheet displaying the narrators' names and faces, and were instructed to refer to the sheet whenever the narrator voice changed.

Each section of the audio guide was broken down into multiple tracks, each associated with an item. We cut the guide into separate tracks rather than using one continuous recording to allow for individual differences in walking speed between items. Each track was initiated by the participant by pressing a button on the MP3 player (NAT) or the spacebar on the computer (LAB). Upon arrival at target items, the guide instructed participants to examine them for 6 s (controlled by a silent period in the recording), which could be preceded or followed by reading information cards associated with target items. Upon completion of this viewing period, the audio guide directed participants to the next target item and instructed them to "press next" when they arrived. Participants in both conditions were given extensive instructions before the tour began, and they were given an opportunity to practice using the MP3 player or the slideshow to control the audio guide. The NAT tour took on average 23.00 min ($SD = 3.02$) for younger adults and 27.15 min ($SD = 4.32$) for older adults. The LAB tour took on average 25.13 min ($SD = 3.89$) for younger adults and 26.83 min ($SD = 4.61$) for older adults (tour duration data were not consistently recorded during the early testing stages, and are missing for 23 out of 80 participants). The experimenter unobtrusively observed participants in both conditions to verify that they followed the instructions. All participants navigated successfully through the tour with little or no experimenter interference.

The NAT and LAB tours were designed under the constraints that they involved the same target content with the same ordinal structure, were controlled by the same audio guide, and that memory for the details common to both tours could be assessed with a single test. These constraints placed some limitations on the degree to which the NAT and LAB conditions were representative of typical naturalistic and laboratory encoding conditions. For instance, we opted to use an audio guide rather than free exploration so as to prioritize the homogeneity of the experience and the comparison between NAT and LAB. On the other hand, we decided to not use a movie or virtual reality for the NAT condition because this would have compromised the physical aspect of the experience. The LAB condition, on the other hand, was designed to capture the timing and perceptual details of the target items in discrete photographs. Thus, the two conditions were matched for content while differing in physical engagement.

2.2.2. The recognition test

We adapted the signal detection logic of old/new recognition tests to a novel true/false test comprising 80 statements (40 true and 40 false) about details from the tour experience (for a similar approach, see Armson et al., 2017). The items are listed in Supplementary Tables 1 and 2. An additional 6 statements were excluded because they pertained to tour details that changed over the course of the study. Lures were created by altering details of tour elements and were non-redundant with respect to true statements (i.e. no two statements referred to the same detail). Most statements referred to item features, rather than whole items, so that there were multiple T/F statements for most target items in the tour (e.g. one statement about the colour of a piece of art and another about its shape). Previous research suggests that distinct details of naturalistic stimuli are forgotten at different rates

(Brady, Konkle, Alvarez, & Oliva, 2013). Statements pertained to four *a priori* categories of information from the tour event: event details (things that happened; e.g. "You passed two fish tanks"; $N = 19$), perceptual details (visual features of target items; e.g. "Al Green's statue called 'The Sage' is green;"; $N = 27$), spatiotemporal order (the location or sequence of items; e.g. "You passed the Hall of Honour before the giant egg art piece"; $N = 16$), and verbal information (content read from information cards or spoken by the audio guide narrators; e.g. "One of the women in the Natalie Ross paintings is named Dora"; $N = 18$). Given that episodic autobiographical memories routinely incorporate many different types of information, we intended to probe memory for diverse aspects of the encoded experience (while ensuring that all tested details were equally accessible in both naturalistic and laboratory encoding conditions).

The subjective quality of participants' memory was assessed using the Remember-Know-Guess paradigm (Tulving, 1985; Gardiner, Ramponi, & Richardson-Klavehn, 2002), wherein Remember responses indicate conscious recollection of contextual details surrounding the encoding of the detail in question, and Know responses indicate a feeling of recognition in the absence of recollection. Know responses are purified by the option to report guessing, which is not a part of the original formulation of the Remember-Know paradigm (Gardiner et al., 2002). Beyond including a Guess option, the present design differs from most previous applications of this method in two ways. First, Remember/Know/Guess responses are typically employed in old/new recognition paradigms where they characterize subjective memory for words or objects as a whole, whereas here these responses are made at the featural level (e.g. a target item's colour). Participants were thus instructed to report "remembering" when they made their true/false responses on the basis of re-experiencing the detail in question (for instance, by seeing the detail in their mind's eye, or by remembering a thought they had about it). Participants were instructed to report "knowing" when they had a feeling, even a strong feeling, that a given statement was true or false without re-experiencing the detail in question. Second, in contrast to old/new paradigms where subjective ratings are only gathered for "old" responses, Remember/Know/Guess responses were gathered for "false" as well as "true" responses, on the basis that our false statements altered features of otherwise valid memoranda and thus would similarly require recollection of the true item feature in question. We used a modified procedure from our prior studies assessing subjective memory judgements in samples of patients with prefrontal cortical pathology (Söderlund, Black, Miller, Freedman, & Levine, 2008; Stamenova et al., 2017), designed to ensure that the participants understood the instructions. We used the terms "Memory Type A" and "Memory Type B" to avoid contamination by pre-experimental associations with the words "remember" and "know". Participants verbally justified their subjective ratings in a set of practice questions, with additional explanation and clarification provided by the examiner (if necessary). Thus, following practice items, all participants demonstrated the capacity to provide justified responses of their subjective mnemonic experience.

For a subset of recognition statements ($N = 55$) participants also made a source memory judgment about which of the four audio guide narrators was associated with the target item in question. These data are not reported here.

2.3. Procedure

Using a between-subjects design, all participants were pseudo-randomly assigned to complete either the naturalistic tour condition (NAT) or the laboratory tour condition (LAB). The study consisted of two sessions. During the first session, participants underwent a neuropsychological assessment battery, including the Rey Auditory Learning Verbal Test (RAVLT), Brief Visuospatial Memory Test (BVM-T-R), face-name associative memory test (Troyer et al., 2012), Symbol Digit Modalities Test (SDMT), Trail Making Test (TMT), phonemic

Table 1

Demographics and neuropsychological test performance split by group (NAT = naturalistic, LAB = laboratory).

Test	Young NAT Mean (SD)	Young LAB Mean (SD)	Old NAT Mean (SD)	Old LAB Mean (SD)
Age (years)	23.82 (3.92)	22.90 (3.58)	69.00 (3.07)	70.89 (4.29)
Education (years)	15.64 (1.09)	15.35 (1.53)	16.50 (3.73)	16.21 (2.23)
Sex (counts; women/men)	14/8	13/7	15/4	13/6
RAVLT: learning trials 1-5	59.27 (7.12)	58.15 (7.15)	53.47 (7.40)	53.26 (6.76)
RAVLT: delayed recall	12.86 (2.01)	12.55 (1.93)	11.47 (2.04)	10.74 (4.03)
RAVLT: recognition hits-FA	14.05 (1.13)	13.55 (1.54)	12.42 (2.69)	11.32 (3.84)
BVMT: total learning	26.68 (4.07)	28.40 (4.26)	19.37 (7.99)	18.68 (6.32)
BVMT: delayed recall	10.59 (1.47)	10.80 (1.11)	7.05 (3.03)	7.57 (2.34)
Face-name associative memory	0.83 (0.16)	0.77 (0.21)	0.64 (0.27)	0.53 (0.27)
SDMT: delayed digit recall ^a	7.32 (2.36)	6.37 (2.61)	3.78 (2.34)	4.63 (2.41)
SDMT: total correct	60.45 (8.24)	59.60 (9.31)	45.42 (10.99)	44.26 (9.19)
TMT: B – A time (s)	28.32 (7.41)	31.47 (16.37)	36.58 (18.09)	41.58 (18.74)
FAS: total correct	43.50 (12.63)	44.60 (10.42)	42.79 (10.46)	42.53 (12.83)
Shipley: total correct	29.52 (4.49)	29.58 (4.52)	35.84 (3.34)	34.92 (3.79)

^a At the end of the neuropsychological test session, participants were presented with each of the symbols and had to recall the corresponding digit.

fluency (FAS), and the Shipley Vocabulary Test. They then completed either the naturalistic tour (NAT) or the laboratory tour (LAB). The second session was two days (48 \pm 3 h) after the first, and comprised the recognition memory test followed by the Autobiographical Interview, a semi-structured interview probing recall for the tour (Levine et al., 2002). Autobiographical Interview data are reported in Diamond, Armson, and Levine (in revision) and Diamond and Levine (in revision), investigating recall accuracy and temporal contextual organization, respectively. Participants were instructed to turn their phone off or on silent during all aspects of testing and leave all their belongings in the locked test room during the tour. All participants exited and re-entered the building through specific doors to limit re-exposure to the tour.

2.4. Analysis

We conducted separate multiple correspondence analyses (MCAs) on the true/false and Remember/Know/Guess responses using the ExPosition and InPosition packages (Beaton, Fatt, & Abdi, 2014) in the R programming language. For a detailed description of MCA, see Abdi and Valentin (2007), and for a recent and similar approach, see Armson et al. (2017). Briefly, MCA is a multivariate data reduction technique, similar to principal components analysis (PCA), but tailored for the analysis of nominal or ordinal data. MCA is ideal for our true/false memory test (1) because MCA identifies patterns of correlation among test responses that explain the maximal amount of variance, and (2) because MCA transforms categorical data into quantitative variables—a feature allowing for more detailed analysis of memory performance. MCA extracts orthogonal dimensions (also called factors or components) that best explain the variance (called ‘inertia’ in this context) of the observations (i.e. participants) and variables (i.e. true/false responses to each item), with each dimension extracting a portion of the total inertia of the data. As in other multivariate data reduction techniques, each observation and variable has a factor score for each dimension representing the importance of this observation or variable for this dimension. These factor scores, in turn, can be used as coordinates in the graphical representation of a dimension space (e.g. a given participant and/or recognition response can be represented as a point in a scatterplot in the 2-D space defined by its Dimensions 1 and 2 factor scores). In correspondence analysis, the inertia explained by a dimension is called the eigenvalue of this dimension (denoted by λ ; it is always smaller than 1), and the percentage of the total explained variance is denoted by τ .

We evaluated the significance of the overall MCA decomposition and of each dimension with a permutation test by which we permuted the participants’ responses independently for each true/false item 1000

times. For each iteration, we computed the total inertia (equivalent to a χ^2 statistic; see Abdi & Béra, 2018) and the eigenvalue associated with each dimension. The proportion of observed inertia and eigenvalues larger than the observed values provides an empirical p -value.

Recall that for each dimension, the true and false (or Remember, Know, and Guess) responses for each item are represented by their factor scores. To measure the reliability of these factor scores, we used a bootstrapping procedure by which 1000 new datasets were generated by resampling the observations with replacement. The ratio of the mean of a factor score to its standard deviation (called a “bootstrap ratio”) can be interpreted like a t -statistic. We analyzed the collection of responses with bootstrap ratios of magnitude of 2 or more—a value roughly corresponding to uncorrected p -values of 0.05 — to interpret the meaning of the dimensions extracted by the MCA analyses. In other words, the responses significantly loading on the positive and negative poles of each dimension together reveal the pattern of memory performance captured by this dimension.

3. Results

3.1. Demographics and neuropsychological test performance

The demographics and neuropsychological test data for the four participant groups are reported in Table 1. First, we conducted a priori t -tests (uncorrected for multiple comparisons) to compare younger and older adults’ neuropsychological test performance. As expected, older adults performed significantly worse than younger adults on all tests (p ’s < .009), except verbal fluency (FAS) where there was no age difference (p = .598), and the Shipley Vocabulary Test where older adults out-performed younger adults (p < .001). Next, to confirm that the encoding conditions were matched in age, education, and neuropsychological test performance, we conducted t -tests (uncorrected for multiple comparisons) to compare naturalistic (NAT) and laboratory (LAB) participants within each age group. Among younger adults, there were not significant condition differences in age ($t(40)$ = 0.79, p = .435), education ($t(40)$ = 0.70, p = .487), nor in any neuropsychological test (all p ’s > .19). Similarly, among older adults, there were not significant condition differences in age ($t(36)$ = 1.56, p = .127), education ($t(36)$ = 0.27, p = .789), nor in any neuropsychological test (all p ’s > .23). No condition difference was observed when collapsing across age groups (all p ’s > .14).

3.2. Objective memory: True/false accuracy

We first conducted MCA on participants’ true/false responses. To reiterate, MCA extracts orthogonal latent dimensions that best explain

the variance across participants and recognition test responses. Both observations (i.e. participants) and variables (i.e. true/false responses to each item) have factor scores reflecting their importance for each dimension. Graphically, the values of the observations or variables for these dimensions can be used as coordinates for a scatterplot. Since the recognition test was identical across the four factorial groups (Younger/Older \times NAT/LAB), we were able to submit all participants' responses to a single MCA, which was blind to the age and event condition groups. This approach allowed us to (1) extract the patterns of recognition responses that best explain the variance in memory performance across participants, and then (2) identify whether the age and encoding groups reliably differed in expression of these retrieval patterns.

The overall MCA decomposition was significant ($p \leq .001$). The permutation test and the scree plot revealed two significant ($p < .05$) dimensions (Fig. S2A; Dimension 1: $\lambda_1 = 0.003$, $p \leq .001$, $\tau_1 = 31\%$; Dimension 2: $\lambda_2 = 0.002$, $p \leq .001$, $\tau_2 = 15\%$, where λ is the eigenvalue and τ is the percentage of inertia, or variance, extracted by a given dimension). We projected individual participants onto the resultant 2-dimensional factor space, where the participants' location was determined by their Dimension 1 (x coordinate) and Dimension 2 (y coordinate) factor scores, and we also plotted the means for each group along with their bootstrap-derived 95% confidence intervals around them (Fig. 2, coloured blobs). Separation between groups in the factor space therefore indicates significant group differences at the $p = .05$ level. To interpret the meaning of each dimension (i.e. which aspects of

memory performance are driving variance along that dimension), we investigated the pattern of recognition responses loading on each dimension, isolating responses with bootstrap ratios of magnitude larger than 2. We used the logic of signal detection theory to interpret these responses, calculating the proportion of hits ("true" | true), misses ("false" | true), correct rejections ("false" | false), and false alarms ("true" | false) at each pole of the two dimensions. These proportions are plotted at the poles of the two dimensions in Fig. 2. All recognition responses and associated bootstrap ratios are presented in Supplementary Tables 1 and 2.

Dimension 1 reflected lure discrimination ability. Responses significantly loading on the negative pole (Fig. 2, left; 28 significant item responses) were predominantly false alarms, and on the positive pole (Fig. 2, right; 18 significant item responses), correct rejections. Dimension 2 reflected response bias, with 'false' responses (predominantly misses) defining the negative pole (Fig. 2, bottom; 15 significant item responses) and 'true' responses (hits and false alarms) at the positive pole (Fig. 2, top; 17 significant item responses).

The four participant groups are ordered sequentially along Dimension 1, with NAT being higher than LAB, and younger adults being higher than older adults. This pattern suggests that both younger age and more naturalistic encoding conditions promote greater discrimination of altered details from true details. Furthermore, younger and older groups overlap along Dimension 1 in LAB but not in NAT—a configuration suggesting a significant effect of age on lure

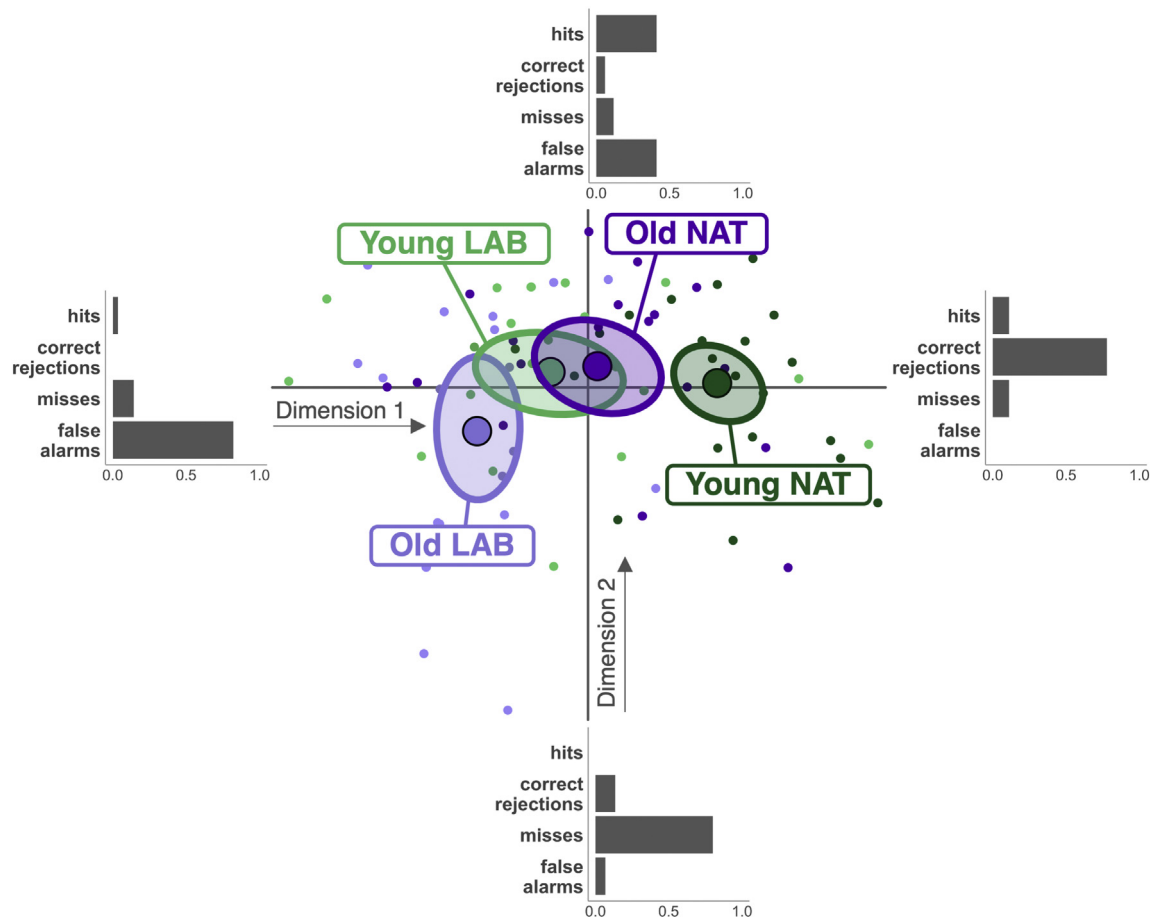


Fig. 2. Results from MCA of true/false responses. The factor space is defined by the first two dimensions, with Dimension 1 as the x-axis and Dimension 2 as the y-axis. Dimensions 1 and 2 explained respectively 31% and 15% of the data variance. Small coloured dots represent individual participants, whose locations in factor space are determined by their Dimensions 1 and 2 factor scores. Group mean coordinates are represented by dark filled circles, and bootstrap-derived 95% confidence intervals are represented by ellipses around the group means. The bar plots at each pole depict the proportion of hits, correct rejections, misses and false alarms among the recognition responses loading significantly on each pole of each dimension. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

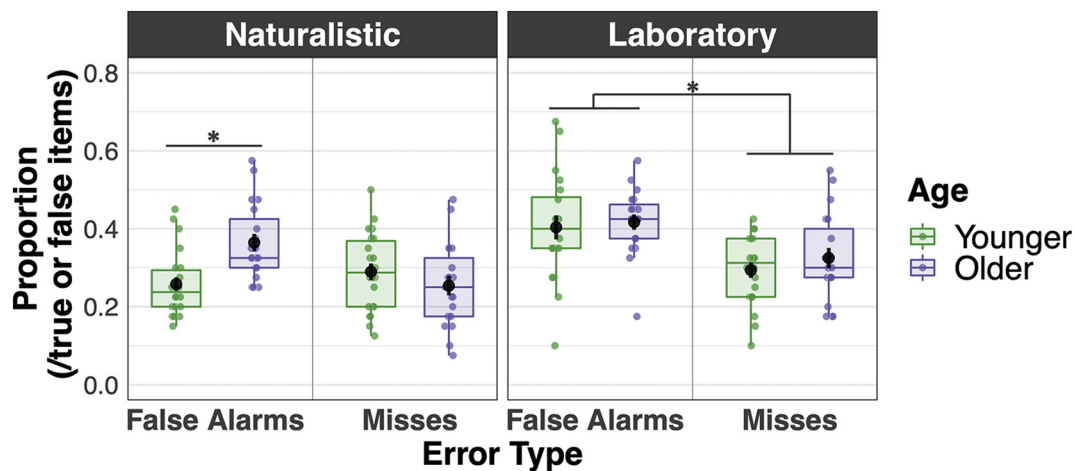


Fig. 3. False alarms as a proportion of false recognition statements ($N = 40$) and misses as a proportion of true recognition statements ($N = 40$). Coloured dots, depicting individual participants, are slightly horizontally jittered to reveal overlap. Black dots depict means and error bars represent standard errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

discrimination in the naturalistic but not in the laboratory encoding conditions.

Younger participants in the naturalistic condition separate from the other groups and have the smallest variance (as shown by the smallest ellipsis along Dimension 1). The specificity of the age-related discrimination impairment to the NAT condition was supported by a 2 (age) \times 2 (condition) \times 2 (error type; false alarms vs. misses) mixed-design ANOVA which indicated an interaction between all three variables ($F(1,76) = 5.38, p = .023$). Follow-up age \times error type ANOVAs within each condition revealed an interaction in the NAT condition ($F(1,39) = 9.86, p = .003$), where older adults committed more false alarms than younger adults ($t(39) = 3.76, p = .001, d = 1.18$), with no age difference in misses ($t(39) = 1.10, p = .554, d = 0.35$; p -values Bonferroni-corrected for two comparisons). There was no such interaction nor age effect in the LAB condition ($F(1,37) = 0.12, p = .735$ and $F(1,37) = 0.94, p = .339$, respectively), but there was a main effect of error type ($F(1,37) = 14.96, p < .001$), with more false alarms than misses overall ($t(38) = 3.91, p < .001, d = 0.94$) (see Fig. 3). Parsed another way, the effect of age on false alarms was specific to the NAT condition ($F(1,76) = 4.04, p = .048$).

To investigate whether true/false responses to items from the four categories (event, perceptual, spatiotemporal, and verbal) loaded differentially on Dimension 1, we conducted one-way ANOVAs separately on positive and negative Dimension 1 response loadings as a function of category. There was not a significant category difference in either direction (positive: $F(3,76) = 1.20, p = .314$; negative: $F(3,76) = 1.42, p = .244$). This finding suggests that the pattern of responses expressed by Dimension 1 was not specific to the type of content retrieved. There was no condition \times category interaction on overall univariate accuracy scores (see Fig. S3).

There was very little differentiation of groups along Dimension 2. Older adults in the LAB condition exhibited the most variance in the negative direction, suggesting a more conservative response bias, although the overlap of ellipses along this dimension suggests that the difference is not statistically reliable.

3.3. Subjective memory: Remember/Know/Guess responses

Next, we ran an MCA on participants' Remember/Know/Guess responses (irrespective of true/false). The overall MCA decomposition was significant ($p = .001$). The permutation test and scree plot revealed four significant ($p < .05$) dimensions (Fig. S2B; Dimension 1: $\lambda_1 = 0.021, p = .001, \tau_1 = 34\%$; Dimension 2: $\lambda_2 = 0.014, p = .001, \tau_2 = 22\%$; Dimension 3: $\lambda_3 = 0.003, p = .001, \tau_3 = 5\%$; Dimension 4:

$\lambda_4 = 0.002, p = .001, \tau_4 = 5\%$). We restricted our interpretation to Dimensions 1 and 2. Dimensions 3 and 4, though significant, explained considerably less variance and were not clearly interpretable. For completeness, the Dimensions 3 vs. 4 factorial map is presented in Fig. S4. As above, we projected group means and bootstrap-derived confidence intervals onto the resultant 2-dimensional factor space (see Fig. 4). The proportion of Remember/Know/Guess responses, out of all responses significantly loading on each dimension, are plotted for each pole of each dimension.

Between-group variance was once again oriented along the first dimension, which was defined on the negative pole by Guess responses (Fig. 4, left; 54 significant item responses) and on the positive pole by Remember responses (Fig. 4, right; 44 significant item responses). The naturalistic encoding groups were higher than the laboratory encoding groups on this dimension, a configuration indicating that naturalistic encoding elicited higher rates of subjective recollection and laboratory encoding elicited higher rates of guessing. Younger participants in each encoding condition overlapped on this dimension, indicating that they did not significantly differ in their subjective response profiles. In contrast, the two older groups exhibited clear separation, with NAT encoding promoting significantly more subjective recollection and less guessing than the LAB condition. Notably, younger and older NAT participants had highly overlapping distributions, indicating similar subjective memory responses. This pattern contrasts with the clear separation between younger and older NAT participants in discrimination accuracy.

As above, there was little differentiation among groups along Dimension 2, which was defined by Know responses. Know responses were absent at the negative pole (Fig. 4, bottom; 39 significant item responses) and constituted 100% of significant responses at the positive pole (Fig. 4B, top; 49 significant item responses). In the LAB condition, associated with less subjective recollection, there is a visual trend for younger participants to report more Know responses, suggesting reliance on familiarity, whereas older adults report more guesses.

Given a priori theoretical interest in the distinction between Remember and Know responses, approximately reflecting recollection and familiarity, respectively (Gardiner et al., 2002; Tulving, 1985), we conducted a 2 (condition) \times 2 (age) \times 2 (response; Remember/Know) ANOVA on subjective response frequency. Consistent with condition differences along subjective Dimension 1, we found a significant interaction between condition and response ($F(1,76) = 5.19, p = .026$), whereby NAT elicited more Remember ($t(78) = 3.64, p < .001, d = 0.81$), but not Know ($t(78) = 0.39, p = .696, d = 0.09$) responses than LAB (Fig. 5).

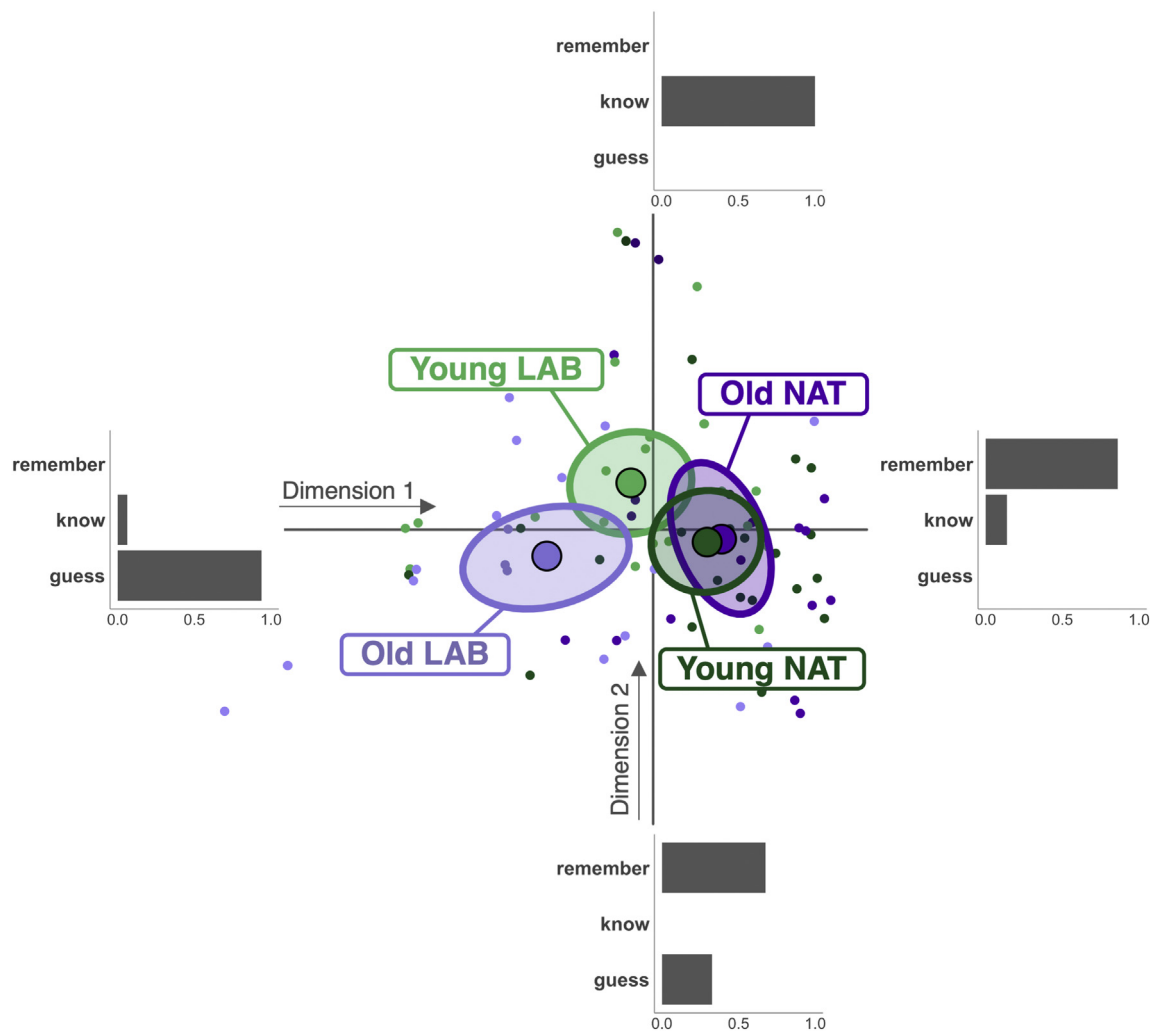


Fig. 4. Results from the MCA of Remember/Know/Guess responses. The factor space is defined by the first two significant dimensions, with Dimension 1 forming the x-axis and Dimension 2 forming the y-axis. Dimension 1 explained 34% of the variance, and Dimension 2 explained 22% of the variance. Group mean coordinates are represented by dark filled circles, and bootstrap-derived 95% confidence intervals are represented by ellipses around the group means. The bar plots at each pole depict the proportion of Remember/Know/Guess responses among the recognition items equal or greater than a bootstrap ratio of 2 at each pole.

3.4. Connecting objective and subjective memory measures

The objective memory MCA revealed that the age and event conditions principally differed along a dimension of objective recollection (discrimination accuracy), and the subjective memory MCA revealed that the age and event conditions principally differed along a dimension

of subjective recollection (self-reported re-experiencing). To investigate the relationship between these measures within participants, we derived objective – subjective Dimension 1 differences scores for each participant (objective Dimension 1 factor score minus Subjective Dimension 1 factor score). We submitted these difference scores to a 2 (age) × 2 (condition) ANOVA (see Fig. 6). Positive scores indicate that

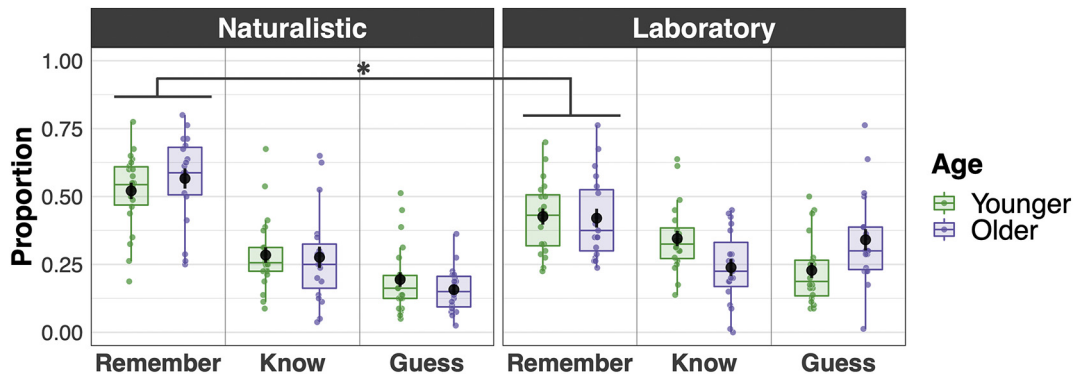


Fig. 5. Remember, Know, and Guess responses as a proportion of total responses in each age group and encoding condition. Note that Guess responses were not analyzed here.

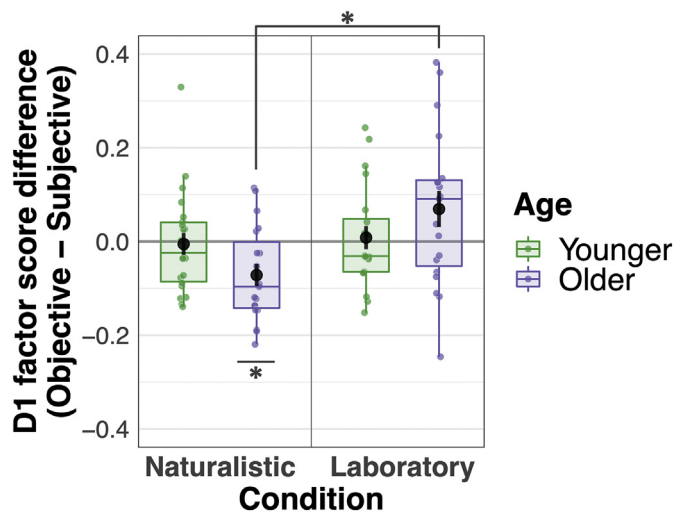


Fig. 6. Difference of objective Dimension 1 participant factor scores minus subjective Dimension 1 participant factor scores within participants. Lower difference scores indicate that subjective re-experiencing exceeded discrimination accuracy, and high difference scores indicate the reverse.

objective exceeded subjective recollection, negative scores indicate the reverse, and scores closer to zero indicate closer alignment between the two. There was a main effect of condition ($F(1,76) = 7.70, p = .007$), in that NAT encoding elicited greater subjective than objective scores (difference $M = -0.036, SD = 0.11$) whereas LAB encoding elicited the reverse (difference $M = 0.038, SD = 0.14; t(78) = 2.61, p = .011, d = 0.58$). This was qualified by an interaction ($F(1,76) = 5.25, p = .025$), where the condition difference was expressed in older ($t(36) = 3.14, p = .003, d = 1.02$) but not in younger adults ($t(40) = 0.40, p = .694, d = 0.12$). Only older NAT participants were significantly different from zero ($t(18) = -3.08, p = .026, d = 0.71$; other p 's > 0.35 , Bonferroni-corrected for 4 comparisons), a pattern suggesting that their objective and subjective memory were uniquely decoupled at the group level, with subjective re-experiencing significantly exceeding discrimination accuracy.

4. Discussion

The assumption that laboratory tests of episodic memory capture the core processes underlying memory for real-world experiences is fundamental to experimental research and clinical practice. Although there is recent neuroimaging evidence challenging this assumption, it is unknown whether and how the content encoded in real-world versus laboratory-style contexts produce different memory representations, while holding constant incidental task features. In this experiment, younger and older adults experienced an event encoded in a naturalistic (active movement through real-world space) or laboratory (passive viewing of discrete photographs) context, followed 48 h later by the same test of memory for specific event details and accompanying subjective memory states. Real-world encoding elicited greater discrimination accuracy and subjective re-experiencing than laboratory encoding. Real-world encoding, however, decoupled objective and subjective recollection in older adults, who reported similar rates of subjective recollection as younger adults despite making significantly more false alarms.

Using a multivariate data-driven analysis, we extracted latent dimensions capturing patterns of responses that explained the most variance in our novel memory test, and found separation of age and encoding groups along these dimensions. Whereas any univariate difference between naturalistic and laboratory encoding conditions could itself be trivial, our results highlight group differences in statistically robust and specific retrieval patterns. The most significant

dimension in objective memory responses captured discrimination of true from altered details (correct rejections versus false alarms). Lure rejection is often based on recollecting veridical item features, especially when lures and targets are highly similar (Gallo, Cotel, Moore, & Schacter, 2007; Migo, Montaldi, Norman, Quamme, & Mayes, 2009). False alarms, on the other hand, are rated as less vivid and contain fewer sensory and contextual details than correctly recognized items (Mather, Henkel, & Johnson, 1997; Norman & Schacter, 1997). Older adults were significantly lower than younger adults on this dimension, extending established evidence of age-related declines in detailed recollection (Levine et al., 2002; Old & Naveh-Benjamin, 2008) and increases in lure susceptibility (Devitt & Schacter, 2016; Diamond et al., 2018; Yassa, Mattfeld, Stark, & Stark, 2011).

Naturalistic encoding, however, elicited greater discrimination accuracy than laboratory encoding for both younger and older adults. This is broadly consistent with well-documented effects of encoding distinctiveness on lure rejection in both age groups (Schacter et al., 1999). Yet contrary to accounts that naturalistic conditions reduce age differences in episodic memory accuracy (e.g. Uttil & Graf, 1993), we found that, if anything, age differences were magnified in the naturalistic condition, where older adults made significantly more false alarms than younger adults. A second dimension pertaining to response bias was significant but showed less sensitivity to age or encoding condition. This overall latent structure is similar to that of a prior study assessing true/false recognition for a staged event at delays of 1 month to 4 years (Armson et al., 2017).

In our memory test, false statements were altered descriptions of item and event features (e.g. changing the colour of a given statue, or the number of times participants were instructed to check the time), analogous to perceptually similar lures in visual mnemonic discrimination tasks (Stark, Kirwan, & Stark, 2019). Accurate responses, therefore, likely required high-fidelity retrieval of specific details. The verbal (rather than visual) nature of our recognition cues likely increases demands on subject-initiated imagery and reconstruction. The demands on fine-grained detail reconstruction may be one reason why we observed a difference between naturalistic and laboratory contexts whereas previous studies testing recognition of photographic stimuli did not (Cabeza et al., 2004).

In the subjective response data, naturalistic encoding again elicited higher factor scores on the first dimension, which was defined by Remember responses on the positive pole and Guesses on the negative pole. Naturalistic encoding thus produced higher rates of self-reported re-experiencing. A follow-up univariate analysis revealed that the encoding manipulation affected Remember but not Know responses. Relative to laboratory stimuli, items encountered in naturalistic environments are bound to a richer and more diverse set of contextual cues — for example, allocentric location, viewpoint, extraneous environmental details, postural and idiothetic information, goal states, multisensory information, etc. More variable contextual associations across items should facilitate more discriminable and richly recollected memories (Schacter et al., 1999; Shiffrin & Steyvers, 1997). Even when stimulus content is held constant, properties characteristic of naturalistic encoding — e.g., volitional versus passive visual exploration (Voss, Gonsalves, Federmeier, Tranel, & Cohen, 2011), active self-motion versus passive displacement (Holmes et al., 2018), self-performed versus observed or imagined actions (Conway & Dewhurst, 1995), and first-person versus observer-based bodily reference frames (Bergouignan, Nyberg, & Ehrsson, 2014) — enhance the formation of episodic associations. More controlled and incremental manipulations may reveal which specific cues and encoding properties drive the different patterns of memory accuracy and subjective quality observed in the present study. It is ultimately these cues and processes that are of interest.

Despite committing significantly more false alarms, older participants' subjective responses overlapped in factor space with those of younger adults in the naturalistic condition. Thus, older adults'

objectively poorer discrimination was not associated with similarly reduced subjective re-experiencing. This interaction held specifically in the naturalistic condition, where older participants' subjective and objective recollection performance were uniquely decoupled. Similar subjective memory inflation, where older adults report equal or greater subjective memory ratings (e.g. confidence, vividness, or reliving) than younger adults in the face of objectively poorer accuracy, has been observed in previous laboratory (Dodson, Bawa, & Krueger, 2007; Jacoby & Rhodes, 2006; McDonough, Cervantes, Gray, & Gallo, 2014) and naturalistic episodic memory studies (Robin & Moscovitch, 2017; St. Jacques, Montgomery, & Schacter, 2015; St-Laurent, Abdi, Bondad, & Buchsbaum, 2014). This pattern of subjective-objective decoupling with age is particularly noteworthy in the present study given that Remember responses are explicitly meant to be based on retrieving specific contextual details.

Why should naturalistic but not laboratory encoding conditions give rise to similar patterns of subjective recollection yet objectively less accurate representation of details in older relative to younger adults? It is important to consider that "Remember" responses permit re-experiencing of any detail associated with the item in question, whereas our true/false statements probed specific item and event features. One possibility is that real-world encoding did indeed support similar subjective recollection in younger and older adults, but that for older adults, the recollected information neither included nor supported retrieval of the specific item features in question (i.e. non-criterial recollection; Yonelinas & Jacoby, 1996). Relative to younger adults, older adults tend to recall relatively fewer objective event features (e.g. perceptual and spatial details), which were the focus of our memory test, and correspondingly more thoughts and feelings (Hashtroudi et al., 1990). Older adults may have committed more false alarms based on recollection of item features not including the detail in question, analogous to a 'recall-to-accept' strategy, which is suboptimal when lures and targets are highly similar (Migo et al., 2009).

On the other hand, subjective-objective decoupling in older adults may reflect a metamemory or retrieval monitoring impairment (Mitchell & Johnson, 2009; Wong, Cramer, & Gallo, 2012). Declines in prefrontal cortical function with age are associated with false recollection, perhaps due to reductions in strategic retrieval and evaluating the veracity of retrieved content (McCabe, Roediger, McDaniel, & Balota, 2009; Mitchell & Johnson, 2009; Trelle, Henson, Green, & Simons, 2017). We found that this effect was specific to the naturalistic condition, whereas older adults' subjective ratings in the laboratory condition were more aligned to their objective memory performance. Given that the retrieval cues were identical across conditions, this interaction suggests that naturalistic versus laboratory encoding conditions produce qualitatively different patterns of episodic memory decline in older age, and potentially in other populations.

4.1. Limitations

We note that our naturalistic encoding condition, like the laboratory condition, was controlled by an audio guide in order to establish homogeneity across participants and correspondence between the naturalistic and laboratory experiences. Our goal was to bring under control low-level event features that are jointly present in autobiographical experiences and absent in typical laboratory encoding conditions (e.g. environmental-scale spatial processing, continuous temporal structure, and volitional control). Although the audio-guided tour reduced volition, participants were in a complex and immersive environment, searching for specific items and physically moving between them. In the laboratory condition, the audio instructions were the same, but participants merely advanced through pre-set photographs. Thus, while laboratory participants did have minimal control over stimulus timing, they could not influence the order, content or viewpoint of the images. Notably, in location-fixed rats, even the potential for exploration elicits place-specific hippocampal activity, and

this specificity is abolished when the potential for exploration is removed by physical restraint (e.g. by snugly wrapping the rat's body in a towel; Foster, Castro, & McNaughton, 1989).

Conversely, the laboratory condition in the present study was itself more naturalistic than typical laboratory stimuli. It was multimodal (visual and auditory), superficially goal-directed, and used full-screen photographs of scenes that depicted a first-person trajectory through a visually rich environment. In this sense, it is similar to recent studies of navigation through virtualized versions of real-world environments (e.g. Brunec et al., 2018). With this in mind, the condition effects on subsequent memory in the present experiment may in fact underestimate differences between typical autobiographical and laboratory encoding conditions.

4.2. Conclusions

We directly compared how information encoded in real-world versus laboratory-style contexts is recollected in younger and older adults. We found that real-world encoding produced more accurate and richly recollected memories than a time- and content-matched laboratory condition in both age groups. Yet for older adults, naturalistic encoding elicited similar patterns of subjective recollection as younger adults despite poorer discrimination. Although laboratory tests of memory provide control and standardization necessary for the accumulation of experimental and theoretical evidence concerning mechanisms, they are not necessarily proxies for real-world memory. This has implications for the understanding of memory operations in real life, particularly for older adults, for whom changes in memory constitute the most frequent cognitive complaint.

CRedit authorship contribution statement

Nicholas B. Diamond: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Visualization. **Hervé Abdi:** Software, Writing - review & editing. **Brian Levine:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

We thank Aggie Bacopulos, Nivethika Jeyakumar, Nicky de Souza, and Robert Amaral for their thorough help with stimulus creation and testing. This research was supported by grants to Brian Levine from the Canadian Institutes of Health Research (Grants MGP-62963 and MOP 367366). N.B.D. was supported by an Ontario Graduate Scholarship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2020.104309>.

References

- Abdi, H., & Béra, M. (2018). Correspondence analysis. In R. Alhajji, & J. Rokne (Eds.). *Encyclopedia of social network analysis and mining* (2nd ed.). New York: Springer.
- Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. In N. Salkind (Vol. Ed.), *Encyclopedia of measurement and statistics*. Vol. 1. Thousand Oaks: Sage.
- Aghajani, Z. M., Acharya, L., Moore, J. J., Cushman, J. D., Vuong, C., & Mehta, M. R. (2015). Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality. *Nature Neuroscience*, 18(1), 121–128. <https://doi.org/10.1038/nn.3884>.
- Armstrong, M. J., Abdi, H., & Levine, B. (2017). Bridging naturalistic and laboratory assessment of memory: The Baycrest mask fit test. *Memory*, 25, 1–10. <https://doi.org/10.1080/09658211.2016.1241281>.
- Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist*, 44(9), 1185–1193. <https://doi.org/10.1037//0003-066X.44.9.1185>.
- Beaton, A. D., Fatt, C. R. C., & Abdi, H. (2014). An ExPosition of multivariate analysis with the singular value decomposition in R. *Computational Statistics & Data Analysis*, 72, 176–189.

- Bergouignan, L., Nyberg, L., & Ehrsson, H. H. (2014). Out-of-body-induced hippocampal amnesia. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1318801111>.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General*, 142(3), 791–808. <https://doi.org/10.1037/a0029649>.
- Brandstatt, K. L., & Voss, J. L. (2014). Age-related impairments in active learning and strategic visual exploration. *Frontiers in Aging Neuroscience*, 6(February), 19. <https://doi.org/10.3389/fnagi.2014.00019>.
- Brune, I. K., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z. X., ... Moscovitch, M. (2018). Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Current Biology*, 1–7. <https://doi.org/10.1016/j.cub.2018.05.016>.
- Cabeza, R., Prince, S. E., Daselaar, S. M., Greenberg, D. L., Budde, M., Dolcos, F., ... Rubin, D. C. (2004). Brain activity during episodic retrieval of autobiographical and laboratory events: An fMRI study using a novel photo paradigm. *Journal of Cognitive Neuroscience*, 16(9), 1583–1594.
- Cabeza, R., & St. Jacques, P. L. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, 11(5), 219–227. <https://doi.org/10.1016/j.tics.2007.02.005>.
- Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, 24(4), 403–416.
- Chen, H.-Y., Gilmore, A. W., Nelson, S. M., & McDermott, K. B. (2017). Are there multiple kinds of episodic memory? An fMRI investigation comparing autobiographical and recognition memory tasks. *The Journal of Neuroscience*, 37(10), <https://doi.org/10.1523/JNEUROSCI.1534-16.2017> (1534–16).
- Chow, T. E., Westphal, A. J., & Rissman, J. (2018). Multi-voxel pattern classification differentiates personally experienced event memories from secondhand event knowledge. *NeuroImage*, 176(March), 110–123. <https://doi.org/10.1016/j.neuroimage.2018.04.024>.
- Conway, M. A. (1991). In defense of everyday memory. *American Psychologist*, 46(1), 19–26. <https://doi.org/10.1037/0003-066X.46.1.19>.
- Conway, M. A., & Dewhurst, S. A. (1995). Remembering, Familiarity, and Source Monitoring. *The Quarterly Journal of Experimental Psychology*, 48A(1), 125–140.
- Craik, F. I. M. (1986). A functional account of age differences in memory. *Human Memory and Cognitive Capabilities: Mechanisms and Performances*, 409–422 (April).
- Dede, A. J. O., Frascino, J. C., Wixted, J. T., & Squire, L. R. (2016). Learning and remembering real-world events after medial temporal lobe damage. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1617025113> (201617025).
- Devitt, A. L., & Schacter, D. L. (2016). False memories with age: Neural and cognitive underpinnings. *Neuropsychologia*, 91, 346–359. <https://doi.org/10.1016/j.neuropsychologia.2016.08.030>.
- Diamond, N. B., Armon, M. J., & Levine, B. (2019). *The truth is out there: Accuracy and detail in recall of verifiable real-world events*. (PsyArxiv).
- Diamond, N. B., & Levine, B. (2019). Linking details to temporal context reinstatement in recall of real-world experiences. *PsyArxiv*, 1–41.
- Diamond, N. B., Romero, K., Jeyakumar, N., & Levine, B. (2018). Age-related decline in item but not spatiotemporal associative memory for a real-world event. *Psychology and Aging*, 33(7), 1079–1092. <https://doi.org/10.1037/pag0000303>.
- Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, 22(1), 122–133. <https://doi.org/10.1037/0882-7974.22.1.122>.
- Foster, T. C., Castro, C. A., & McNaughton, B. L. (1989). Spatial Selectivity of Rat Hippocampal Neurons: Dependence on Preparedness for Movement. *Science*, 244, 1580–1582.
- Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based retrieval monitoring: The importance of event distinctiveness. *Psychology and Aging*, 22(1), 209–213. <https://doi.org/10.1037/0882-7974.22.1.209>.
- Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: a meta-analysis of remember, know, and guess responses. *Memory*, 10(2), 83–98.
- Gilboa, A. (2004). Autobiographical and episodic memory—One and the same? Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia*, 42(10), 1336–1349. <https://doi.org/10.1016/j.neuropsychologia.2004.02.014>.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York, N.Y.: Academic Press. Inc.
- Hashtroudi, S., Johnson, M., & Chrosniak, L. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging*, 119–136.
- Hashtroudi, S., Parker, E. S., Luis, J. D., & Reisen, C. A. (1989). Generation and elaboration in older adults. *Experimental Aging Research*, 15(1–2), 73–78. <https://doi.org/10.1080/03610738908259760>.
- Hay, J. F., & Jacoby, L. L. (1999). Separating habit and recollection in young and older adults: Effects of elaborative processing and distinctiveness. *Psychology and Aging*, 14(1), 122–134.
- Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, 34(2), 151–176. <https://doi.org/10.1016/j.intell.2005.09.005>.
- Holmes, C. A., Newcombe, N. S., & Shipley, T. F. (2018). Move to learn: Integrating spatial information from multiple viewpoints. *Cognition*, 178(April), 7–25. <https://doi.org/10.1016/j.cognition.2018.05.003>.
- Jacoby, L. L., & Rhodes, M. G. (2006). False remembering in the aged. *Current Directions in Psychological Science*, 15(2), 49–53. <https://doi.org/10.1111/j.0963-7214.2006.00405.x>.
- Jeunehomme, O., & D'Argembeau, A. (2018). The time to remember: Temporal compression and duration judgments in memory for real-life events. *Quarterly Journal of Experimental Psychology*, 1. <https://doi.org/10.1177/1747021818773082> (1747021818773082).
- Kim, H. (2015). Default network activation during episodic and semantic memory retrieval: A selective meta-analytic comparison. *Neuropsychologia*, 80, 35–46. <https://doi.org/10.1016/j.neuropsychologia.2015.11.006>.
- Koriat, A., & Goldsmith, M. (1996). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences*, 19(02), 167. <https://doi.org/10.1017/S0140525X00042114>.
- Kvavilashvili, L., & Ellis, J. (2004). Ecological validity and the real-life/laboratory controversy in memory research: A critical (and historical) review. *History and Philosophy of Psychology*, 6(1), 59–80.
- Leal, S. L., & Yassa, M. A. (2015). Neurocognitive aging and the hippocampus across species. *Trends in Neurosciences*, 38(12), 800–812. <https://doi.org/10.1016/j.tins.2015.10.003>.
- Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging*, 17(4), 677–689. <https://doi.org/10.1037/0882-7974.17.4.677>.
- Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory and Cognition*, 25(6), 826–837. <https://doi.org/10.3758/BF03211327>.
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., & Balota, D. A. (2009). Aging reduces veridical remembering but increases false remembering: Neuropsychological test correlates of remember-know judgments. *Neuropsychologia*, 47(11), 2164–2173. <https://doi.org/10.1016/j.neuropsychologia.2008.11.025>.
- McDermott, K. B., Szpunar, K. K., & Christ, S. E. (2009). Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, 47(11), 2290–2298. <https://doi.org/10.1016/j.neuropsychologia.2008.12.025>.
- McDonough, I. M., Cervantes, S. N., Gray, S. J., & Gallo, D. A. (2014). Memory's aging echo: Age-related decline in neural reactivation of perceptual details during recollection. *NeuroImage*, 98, 346–358. <https://doi.org/10.1016/j.neuroimage.2014.05.012>.
- 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*, 62(6), 1198–1215. <https://doi.org/10.1080/17470210802391599>.
- Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, 135(4), 638–677. <https://doi.org/10.1037/a0015849>.
- Monge, Z. A., Wing, E. A., Stokes, J., & Cabeza, R. (2018). Search and recovery of autobiographical and laboratory memories: Shared and distinct neural components. *Neuropsychologia*, 110(July 2017), 44–54. <https://doi.org/10.1016/j.neuropsychologia.2017.07.030>.
- Neisser, U. (1978). Memory: What are the important questions. *Practical Aspects of Memory*, 3–19.
- Nilsson, L.-G., Backman, L., Erngrund, K., Nyberg, L., Adolfsson, R., Bucht, G., Winblad, B., et al. (1997). The betula prospective cohort study: Memory, health, and aging. *Aging, Neuropsychology, and Cognition*, 4(1), 1–32. <https://doi.org/10.1080/13825589708256633>.
- Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, 25(6), 838–848. <https://doi.org/10.3758/bf03211328>.
- 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(1), 104–118. <https://doi.org/10.1037/0882-7974.23.1.104>.
- Park, D., Cherry, K., Smith, A., & Lafronza, V. (1990). Effects of distinctive context on memory for objects and their locations in young and elderly adults. *Psychology and Aging*, 5(2), 250–255.
- Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging*, 7(2), 290–298. <https://doi.org/10.1037/0882-7974.7.2.290>.
- Rissman, J., Chow, T. E., Reggente, N., & Wagner, A. D. (2016). Decoding fMRI signatures of real-world autobiographical memory retrieval. *Journal of Cognitive Neuroscience*, 1–10. <https://doi.org/10.1162/jocn>.
- Robin, J. (2018). Spatial scaffold effects in event memory and imagination. *Wiley Interdisciplinary Reviews: Cognitive Science*(August 2017), e1462. <https://doi.org/10.1002/wcs.1462>.
- Robin, J., & Moscovitch, M. (2017). Familiar real-world spatial cues provide memory benefits in older and younger adults. *Psychology and Aging*. <https://doi.org/10.1037/pag0000162>.
- Roediger, H. L., & McDermott, K. B. (2013). Two types of event memory. *Proceedings of the National Academy of Sciences of the United States of America*, 110(52), 20856–20857.
- Rosa, N. M., & Gutches, A. H. (2011). Source memory for action in young and older adults: Self vs. close or unknown others. *Psychology and Aging*, 26(3), 625–630. <https://doi.org/10.1037/a0022827>.Source.
- Rugg, M. D., & Vilberg, K. L. (2012). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, 23, 1–6. <https://doi.org/10.1016/j.conb.2012.11.005>.
- Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, 40(1), 1–24. <https://doi.org/10.1006/jmla.1998.2611>.
- Sharps, M. J., & Gollin, E. S. (1987). Memory for object locations in younger and elderly

- adults. *Journal of Gerontology*, 42(3), 336–341.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving efficiently from memory. *Psychonomic Bulletin & Review*, 4(2), 145–166.
- Söderlund, H., Black, S. E., Miller, B. L., Freedman, M., & Levine, B. (2008). Episodic memory and regional atrophy in frontotemporal lobar degeneration. *Neuropsychologia*, 46(1), 127–136. <https://doi.org/10.1016/j.neuropsychologia.2007.08.003>.
- St. Jacques, P. L., Montgomery, D., & Schacter, D. L. (2015). Modifying memory for a museum tour in older adults: Reactivation-related updating that enhances and distorts memory is reduced in ageing. *Memory*, 23(6), 876–887. <https://doi.org/10.1080/09658211.2014.933241>.
- St. Jacques, P. L., & Schacter, D. L. (2013). Modifying memory: Selectively enhancing and updating personal memories for a museum tour by reactivating them. *Psychological Science*, 24(4), 537–543. <https://doi.org/10.1177/0956797612457377>.
- Stamenova, V., Gao, F., Black, S. E., Schwartz, M. L., Kovacevic, N., Alexander, M. P., & Levine, B. (2017). The effect of focal cortical frontal and posterior lesions on recollection and familiarity in recognition memory. *Cortex*, 91, 316–326. <https://doi.org/10.1016/j.cortex.2017.04.003>.
- Stark, S. M., Kirwan, C. B., & Stark, C. E. L. (2019). Mnemonic similarity task: A tool for assessing hippocampal integrity. *Trends in Cognitive Sciences*, 1–14. <https://doi.org/10.1016/j.tics.2019.08.003>.
- St-Laurent, M., Abdi, H., Bondad, A., & Buchsbaum, B. R. (2014). Memory reactivation in healthy aging: Evidence of stimulus-specific dedifferentiation. *Journal of Neuroscience*, 34(12), 4175–4186. <https://doi.org/10.1523/JNEUROSCI.3054-13.2014>.
- Summerfield, J. J., Hassabis, D., & Maguire, E. A. (2009). Cortical midline involvement in autobiographical memory. *NeuroImage*, 44(3), 1188–1200. <https://doi.org/10.1016/j.neuroimage.2008.09.033>.
- Thome, A., Marrone, D. F., Ellmore, T. M., Chawla, M. K., Lipa, P., Ramirez-Amaya, V., ... Barnes, C. A. (2017). Evidence for an evolutionarily conserved memory coding scheme in the mammalian hippocampus. *The Journal of Neuroscience*, 37(10), <https://doi.org/10.1523/JNEUROSCI.3057-16.2017> (3057–16).
- Trelle, A. N., Henson, R. N., Green, D. A. E., & Simons, J. S. (2017). Declines in representational quality and strategic retrieval processes contribute to age-related increases in false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(12), 1883–1897.
- Troyer, A. K., et al. (2012). Associative recognition in mild cognitive impairment: relationship to hippocampal volume and apolipoprotein E. *Neuropsychologia*, 50(14), 3721–3728.
- Tulving, E. (1983). *Elements of episodic memory*. New York, N.Y.: Oxford University Press.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, 26(1), 1–12.
- Tulving, E. (1991). Memory research is not a zero-sum game. *American Psychologist*, 46(1), 41–42.
- Uttl, B., & Graf, P. (1993). Episodic spatial memory in adulthood. *Psychology and Aging*, 8(2), 257–273. <https://doi.org/10.1037/0882-7974.8.2.257>.
- Voss, J. L., Gonsalves, B. D., Federmeier, K. D., Tranel, D., & Cohen, N. J. (2011). Hippocampal brain-network coordination during volitional exploratory behavior enhances learning. *Nature Neuroscience*, 14(1), 115–120. <https://doi.org/10.1038/nn.2693>.
- Wong, J. T., Cramer, S. J., & Gallo, D. A. (2012). Age-related reduction of the confidence – Accuracy relationship in episodic memory: Effects of recollection quality and retrieval monitoring. *Psychology and Aging*, 27(4), 1053–1065. <https://doi.org/10.1037/a0027686>.
- Yassa, M. A., Mattfeld, A. T., Stark, S. M., & Stark, C. E. L. (2011). Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences*, 1–6. <https://doi.org/10.1073/pnas.1101567108>.
- Yonelinas, A. P., & Jacoby, L. L. (1996). Noncriterial recollection: Familiarity as automatic, irrelevant recollection. *Consciousness and Cognition*, 5(1–2), 131–141. <https://doi.org/10.1006/ccog.1996.0008>.