**Title:** Taking stock of the past: A psychometric evaluation of the Autobiographical Interview

**Short title:** Psychometric assessment of the AI

**Authors:** Amber W. Lockrow[1], Roni Setton[1], Karen A.P. Spreng[2], Signy Sheldon[3], Gary R. Turner[4], R. Nathan Spreng[1,3,5,6]

**Affiliations:**
1 Montréal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada
2 Concordia University, Montréal, QC, Canada
3 Department of Psychology, McGill University, Montréal, QC, Canada
4 Department of Psychology, York University, Toronto, ON, Canada
5 Department of Psychiatry, McGill University, Montréal, QC, Canada
6 McConnell Brain Imaging Centre, Montréal Neurological Institute, McGill University, Montréal, QC, Canada

**Correspondence to:**

R. Nathan Spreng
Laboratory of Brain and Cognition
Montreal Neurological Institute
Department of Neurology & Neurosurgery
McGill University
Montreal, Quebec H3A 2B4
Canada
phone: 514-398-7268
email: nathan.spreng@gmail.com

**Abstract (188/200)**

Autobiographical memory (AM) involves a rich phenomenological re-experiencing of a spatio-temporal event from the past, which is challenging to objectively quantify. The Autobiographical Interview (AI; Levine et *al.*, 2002, *Psychology & Aging*) is a manualized performance-based assessment designed to quantify episodic (internal) and semantic (external) features of recalled and verbally conveyed prior experiences. The AI has been widely adopted yet has not undergone a comprehensive psychometric validation. We investigated the reliability, validity, association to individual differences measures, and factor structure in healthy younger and older adults (N=352). Evidence for the AI's reliability was strong: the subjective scoring protocol showed high inter-rater reliability and previously identified age effects were replicated. Internal consistency across timepoints was robust, suggesting stability in recollection. Central to our validation, internal AI scores were positively correlated with standard, performance-based measures of episodic memory, demonstrating convergent validity. The two-factor structure for the AI was not well-supported by confirmatory factor analysis. Adjusting internal and external detail scores for the number of words spoken (detail density) improved trait estimation of AM performance. Overall, the AI demonstrated sound psychometric properties for inquiry into the qualities of autobiographical remembering.

Autobiographical memory (AM) is a multifaceted form of explicit memory for personal life experiences (Conway & Pleydell-Pearce, 2000; Moscovitch et al., 2005; Rubin, 1988). AM retrieval is characterized by the phenomenological recollection, or re-experiencing, of a prior personal event including its content and spatiotemporal context. This process of re-experiencing situates aspects of AM within the broader domain of episodic memory. AM also involves accessing information from semantic memory. This information includes knowledge surrounding the event, which is not specific to a particular episode, but reflects a generalized and personal knowledge of concepts, facts and meaning (Renoult et al., 2012; Irish et al., 2013, 2012). The distinction between different types of information is necessary to accurately measure the episodicity of recollection yet distinguishing between episodic versus semantic features of naturalistic recollection presents unique measurement challenges.

The Autobiographical Interview (AI) was introduced by Levine and colleagues (2002) to quantify and dissociate episodic and semantic features of AM recall. This semi-structured interview elicits a verbal recounting of specific personal events, sampled across the lifespan. These narratives are transcribed and scored to characterize informational units, or details, within each memory. Details specific to the event, place, and perceptual information are identified as “internal details.” Internal details reflect qualities of the recollective experience and are considered a metric of episodic memory. Details not specific to the event, including broader conceptual and personal information that fall within the domain of semantic memory are coded as “external details”, in addition to repetitions and off-topic comments

In the first published study involving the AI, younger adults showed a bias towards reporting more internal episodic details whereas older adults showed a bias towards reporting more semantic details (Levine et al., 2002). This pattern converged with well-established age-related effects of reduced episodic memory and greater semantic knowledge in older adults (Park et al. 2001; Spreng & Turner, 2019). Additionally, the authors demonstrated convergent validity with the Autobiographical Memory Interview (AMI), a coarser, single factor measure of AM, which rates verbal event recall on a continuum from general knowledge to episodic detail (Kopelman et al., 1989, 1990).

The AI has been widely adopted to characterize the involvement of episodic memory during AM in healthy participants and patient populations (see https://levinelab.weebly.com/ai-testing.html; Miloyan et al., 2019 for review). Consistent with the original report, healthy younger adults provide more internal details and fewer external details than older adults when recalling past personal experiences (Addis et al., 2008; Addis et al., 2010; De Brigard et al., 2016; De Brigard et al., 2017; Ford et al., 2014; Gaesser et al., 2011; Madore et al., 2014; Peters et al., 2019; Robin & Moscovitch, 2017; Spreng et al., 2018; St Jacques & Levine, 2007; Vandermorris et al., 2013; Zavagnin et al., 2016). Patients with medial temporal lobe lesions and episodic memory impairment provide fewer internal details compared to controls (Dede, Franscino et al., 2016; Dede, Wixted et al., 2016, Gilboa et al., 2006; Hilverman et al., 2016; Kirwan et al., 2008; Kwan et al., 2010; Kwan et al., 2015; Kwan et al., 2016; Miller et al., 2020; Race et al., 2011; Rosenbaum et al., 2004; Rosenbaum et al., 2008; Rosenbaum et al.,

2009; Squire et al., 2010; Steinvorth et al., 2005). Individuals with Alzheimer's disease and amnestic mild cognitive impairment, also produce fewer internal details compared with controls (Addis et al., 2009; Bastin et al., 2013; Coelho et al., 2019; Gamboz et al., 2010; Meulenbroek et al., 2010; Murphy et al., 2008; Sheldon et al., 2015). These group comparisons provide converging evidence to support the validity and reliability of the AI as a measure of AM. Few studies have examined individual differences on the AI, and there is heterogeneity in how AM is characterized to map onto reliable person-level factors (e.g., Palombo et al., 2018). To date, a psychometric evaluation of the AI, and how these compare to standard laboratory measures of memory and cognitive function, have not been comprehensively examined.

In the present study we provide a psychometric assessment of the AI in a large sample of healthy younger and older adults. Our aims were to (i) examine the reliability and validity of the AI, (ii) assess relationships with individual difference measures of psychological and cognitive function and (iii) test whether a two-factor solution, capturing the distinction between internal and external event details, accurately reflects the data structure. To address these aims we derived eight metrics from the AI. The first three are consistent with the original protocol (Levine et al, 2002): Number of internal details (internal count), number of external details (external count), and the proportion of internal to total details (proportion score). We also derived two additional measures. Internal and external density which divide detail counts by the total number of words spoken (Spreng et al., 2018). Semantic details and density were examined as more direct measures of semantic information. Number of words spoken was also included as a distinct variable of interest. Thus, a fourth aim was to examine whether controlling for verbosity would impact the psychometric properties of the AI.

To address these aims we conducted five sets of analyses. First, we evaluated the reliability of the AI, examining inter-rater reliability as well as internal consistency by testing stability across event memories, associations among AI scores, and associations with participant ratings (e.g., vividness, personal relevance, emotionality, rehearsal frequency) as well as scorer summary ratings for each event. Next, we evaluated how AI scores varied by demographic factors across our sample, including age, gender and education level. Third, we examined convergent validity by testing associations with laboratory measures of memory and cognition. We then tested associations between the AI and factors previously implicated in AM including depression, decision-making, social cognition and personality. Finally, we conducted a confirmatory factor analysis (CFA) to evaluate the two-factor, internal/external model of the AI. We propose specific predictions for each of the analyses in the corresponding results sections below.

## Methods

### Participants

Participants were recruited from greater Ithaca, NY, US and Toronto, Ontario, Canada and were screened to exclude psychiatric, neurological, or other illnesses that could impair cognitive functioning. 203 younger adults and 158 older adults completed all primary measures of interest. Two younger adults and four older adults were excluded for having scores below 27/30 on the Mini Mental Status Exam (Folstein, Folstein, & McHugh, 1975) combined with fluid cognition scores below a national percentile of 25% on the NIH Cognition Toolbox (Weintraub et al., 2014). Two older adults were excluded for scores above 20/30 on the Geriatric Depression Scale (Sheikh & Yesavage, 1986), indicative of moderate or severe depressive symptoms. One older adult was excluded due to extensive confabulatory statements post-experimental session. The final sample included 201 younger adults (114 women, *M*= 22.4 years; *SD*= 3.28 years; range= 18-34 years) and 151 older adults (82 women, *M* = 68.8 years; *SD*= 6.67; range= 60-92 years). Demographic information for this sample is shown in Table 1.

### Procedure

Participants completed the measures described below over several testing sessions as part of a comprehensive behavioral assessment protocol examining goal-directed behavior.

*Autobiographical Interview*. All participants completed the AI as specified in the original Levine et al. (2002) protocol. Older adults provided detailed descriptions of one event from five different time periods: childhood (up to age 11), teenage years (between age 11 and 18), early adulthood (between age 18 and 30), middle adulthood (between age 30 and 55), and late adulthood (within the previous year). Younger adults provided detailed descriptions of events from three periods: childhood, teenage years, and younger adulthood. For each period, participants were asked to describe an event tied to a specific time and place. Recall for the event was examined at three probe levels: free recall (uninterrupted description of the memory), general probe (general questions to elicit further details), and specific probe (specific, targeted questions to elicit details from different categories). As the original study (Levine et al., 2002) reported no differences in the pattern of results for free recall and general probe, we collapse these probe conditions for each of the metrics. Specific probe was not examined in order to focus on spontaneous, uncued recollection. After recalling each event, participants rated the vividness, emotional change, significance (then and now), and rehearsal frequency on a 6-point Likert scale.

| | Gender | | Age | | | Education | | |
|---|---|---|---|---|---|---|---|---|
| | F | M | *M* | *SD* | Range | *M* | *SD* | Range |
| Younger Adults | 114 | 87 | 22.4 | 3.28 | 18-34 | 15.0 | 1.78 | 12-21 |
| Older Adults | 82 | 69 | 68.8 | 6.67 | 60-92 | 17.4 | 3.04 | 12-26 |
| All | 196 | 156 | | | 18-92 | 16.0 | 2.67 | 12-26 |

*Table 1.* Participant Demographics. Gender distributions, age in years, and education level in years are reported. For the full sample, mean and standard deviation are not included for age given its bimodal distribution. F = Women, M = Men, *M* = Mean, *SD* = Standard Deviation

Participant interviews were audio recorded, anonymized, and transcribed. Each transcription was quality checked with the audio recording to ensure accuracy. All interviews were double-scored by research assistants who were trained on the protocol provided by Dr. Brian Levine. For each memory, scorers identified individual informational units, or details. Details were categorized as either 'internal' or 'external'. Internal details were those having to do with the identified event and were specific in time and place. External details were those that did not include information about the specified event, were not specific in time or place, or consisted of semantic information (i.e., conveyed facts that temporally extended beyond the event). Both internal and external details could be broken down into sub-categories of event, place, time, perceptual, and emotion/thought details. External details also included semantic, repetition, or other details. Detail counts were then tallied. As part of the scoring protocol, scorers also provided summary ratings indicating the level of detail provided. Scorers assigned an overall score for each memory based on ratings for place, time, perceptual and emotion/thought information, time integration, episodic richness, and a global rating of detail and specificity (consistent with the Autobiographical Memory Interview (AMI) score; Kopelman et al., 1990).

*AI Metrics.* Eight dependent variables were calculated from participant interviews:

1. Internal count
2. External count
3. Semantic count
4. Proportion score
5. Internal density
6. External density
7. Semantic density
8. Word count

Internal count (1), external count (2), and the proportion of internal to total details (4) were calculated following the procedure reported in Levine et al. (2002). To account for variation in verbosity, detail counts were divided by the word count for each memory, resulting in measures of internal (5) and external detail density (6; Spreng et al., 2018). To separate personal semantic information from non-mnemonic informational units, we also directly examined semantic details, via count (3) and density (7). We also evaluated word count as a separate variable of interest (8). All dependent variables were calculated for each memory. Composite scores were calculated by averaging across all memories. In addition to these primary AI metrics, we also examined average self-report and scorer ratings.

*Laboratory measures of Episodic Memory, Semantic Memory, and Executive Function*
Index scores of episodic memory, semantic memory, and executive function were derived from the NIH Cognition Toolbox and in-lab tasks. Measures of episodic memory included: Verbal Paired Associates (Wechsler, 2008), the Associative Recall Paradigm (Brainerd & Pressley, 2013), the NIH Rey Auditory Verbal Learning Test (Weintraub et al., 2014), and the NIH Picture Sequence Memory task. Measures of semantic memory

included: the Shipley-2 Test of Vocabulary (Shipley et al., 2009), the NIH Reading Recognition Task, and the NIH Picture Vocabulary Task. Measures of executive function included: NIH List Sorting task, the NIH Card Change Sort Task, the NIH Flanker Task, a Reading Span task (Daneman & Carpenter, 1980), and the Trail Making Task (Reitan, 1958). Index scores were calculated by z-scoring each measure and averaging across measures within each domain. Two younger participants' semantic index scores were winsorized for outlying performance.

A subsample of 148 younger adults and 90 older adults had complete data for the Remember/Know paradigm (R/K; Tulving, 1972; Tulving 1985). Participants were excluded if they provided Remember responses for more than 90% of recognized items, which limited the number of familiarity trials (e.g., Stamenova et al., 2017). Fifty-two participants were excluded on this basis. Recollection and familiarity scores were derived according to standardized methods (Söderlund et al., 2008).

*Measures of depression, decision-making, social cognition and personality*

A subsample of 67 younger adults and 93 older adults completed a measure of temporal discounting (Loeckenhoff et al., 2011). Temporal discounting was evaluated with a computerized forced choice task in which participants selected whether they would prefer to receive small magnitude rewards now or at varied dates in the future (7-180 days). Dependent variables included area under the curve (Myerson et al., 2001), reward index (Boettiger et al., 2007), and the proportion of patient (delayed) choices. To facilitate comparisons with prior work associating the AI with temporal discounting (Lempert, MacNear, et al., 2020), we computed an additional "perceptual/gist ratio" (computed as [(internal time count + internal place count + internal perceptual count) / total internal count]) from the AI scores.

A subset of participants completed additional online self-report questionnaires on Qualtrics. Fifteen younger adults and three older adults were excluded based on failed attention checks. Thus, the maximum sample for analyses that included self-report questionnaires was 162 younger adults and 125 older adults. Depressive symptoms were measured with the Beck Depression Inventory (Beck et al., 1996) for younger adults and the Geriatric Depression Scale (Sheikh & Yesavage, 1986) for older adults. Social Cognition measures included the Interpersonal Reactivity Index (Davis, 1983), the Toronto Empathy Questionnaire (Spreng et al., 2009), and the Reading the Mind in the Eyes Task (Baron-Cohen et al., 1997). Personality was evaluated with the Big Five Aspect Scales (DeYoung et al., 2007).

## Analysis

All analyses were conducted with the R statistical software version 4.0.0 (R Core Team, 2020). The packages used included corrplot (Wei & Simko, 2017), ggplot2 (Villanueva & Chen, 2019), lme4 (Bates et al., 2014), lmerTest (Kuznetsova et al., 2017), lavaan 0.6-7 (Rosseel, 2012), and psych (Revelle, 2012). In order to maximize statistical power, our primary analyses leveraged the entire sample of participants where appropriate, controlling for age group and gender. We also conducted parallel analyses within the

two age groups controlling for gender. Results are summarized in the main text and depicted in Supplementary Figures.

Potential outliers on AI metrics were evaluated using the median absolute deviation method (Leys et al., 2013). In total, 131 observations were winsorized prior to analysis. 37 participants' data contained one outlier, and 18 had more than one. Outlier correction did not change the pattern of results reported here.

Throughout the manuscript, associations were examined among AI metrics (e.g. internal and external count) and between specific AI metrics and other measures with hypothesis-driven relationships (e.g., internal count and episodic memory). For completeness and transparency, relationships between potentially unpredicted associations are also reported. In order to assess validity and to guide future work, uncorrected p-values are reported (see Rothman, 1990; Saville, 1990; Althouse, 2016). Predicted associations are indicated in both the figures and the main text. Significant non-predicted associations are shown in figures and additionally flagged if they survived Bonferroni correction ($\alpha = .05$) based on the number of unpredicted correlations conducted within each set of analyses. All correlation magnitudes are reported in figures to provide estimation of effect sizes based upon the current sample.

(i) Reliability of the Autobiographical Interview

*Inter-rater reliability*. Scoring of autobiographical events is manualized and involves significant training to a set reliability criteria. Scorers must identify the primary event in the narrative, demarcate details, and determine whether these details are internal or external to the event. This decision process is subjective. To evaluate the AI's inter-rater reliability, intraclass correlations were computed based on a mean-rating ($k$=2), absolute agreement, one-way random effects model (Shrout & Fleiss, 1979) between scorers. This was done for composite scores of total detail count, internal count, and external count, and for each of the sub-categories. Correlations were computed on the full sample and separately within younger and older adults. Detail counts were averaged between scorers for all subsequent analyses.

To evaluate the AI's internal consistency, we tested the temporal stability of detail generation with all eight AI variables and subsequently examined the association between the eight AI composite scores averaged across events.

*Internal consistency (temporal stability)*. AMs are sampled from multiple events over the lifespan. AI scores are often derived by taking the average of internal and external details from multiple life events, even though remoteness has profound effects on memory (e.g. Linton, 1975; Rubin & Wenzel, 1996; Wagenaar, 1986). We first examined how detail recollection differed across time periods. We tested for detail differences between the most recent (proximal) and remote (childhood) memories with pairwise t-tests. Then, repeated-measure ANOVAs within age groups were conducted to test for detail differences across all three memories in younger adults and all five memories in older adults. Post-hoc Bonferroni corrected pairwise t-tests ($\alpha = .05$) were conducted between

all pairs of events and AI measures. Next, we quantified how similarly details were recalled across time periods with product-moment correlations. Specifically, we computed correlations between 1) the most remote and the most recent event; 2) each event and the average of all events (akin to a modified item-total correlation); 3) each event and the average of the most recent and remote events; and 4) the average of all events and the average of the first and last events. Correlations were conducted for each of the eight AI measures of interest. As younger and older adults described a different number of memories, the full sample was used for correlations between the most recent and remote events, and separate age groups for the remainder.

*Internal consistency (AI metrics)*. We next examined associations among the eight AI composite measures (i.e., averaged across all memories). Partial product-moment correlations were computed between the AI measures across all participants controlling for age group and gender. Analyses were also repeated within age groups (see Supplemental Material).

Recent work suggests that correlating the average of internal and external counts across events obscures a differential ratio of internal and external detail counts within events (Devitt et al., 2017). Given this possibility, we conducted a hierarchical linear model analysis to determine whether internal count predicted external count. An internal count by age group interaction was modeled as a fixed effect, with the memories from different time periods as intercepts and participants as random effects. We compared different model structures to determine the best fitting model (Judd et al., 2012). Our first model included random intercepts for memory and participant. Our second model also allowed internal count to vary by event. As a chi-square test indicated that fit did not differ between the models ($\chi 2(2) = .24$, $p = .89$), we report results from the simpler random intercept model.

*Internal consistency (AI ratings).* In our final assessment of internal consistency, we examined how the eight primary, performance-based AI measures were related to self-reported participant ratings and scorer summary ratings. Associations were examined with partial Spearman correlations (due the non-continuous nature of Likert ratings) controlling for age and gender.

(ii) Demographic Associations with the Autobiographical Interview

*Autobiographical Interview Performance by Gender, Age and Education*. To determine gender and age differences, 2-way Analyses of Variance (ANOVAs) were conducted between gender and age groups for all eight AI measures of interest. Post-hoc pairwise t-tests with Bonferroni correction were run ($\alpha = .05$) to examine pairwise differences between groups. We also conducted partial product-moment correlations to examine the relationship between education and the eight AI measures across all participants, controlling for gender and age. Correlations were repeated within age groups controlling for gender.

(iii) Convergent Validity of the Autobiographical Interview

Central to validation of the AI is determining whether it shares variance with other validated measures of related constructs. To evaluate convergent validity of the AI, we conducted partial product-moment correlations between the AI measures and index scores of episodic memory, semantic memory, and executive function. We additionally examined associations between the AI measures and recollection and familiarity on the R/K task. In order to maximize statistical power, our primary analyses leveraged the entire sample of participants, controlling for age and gender. Complementary analyses in separate groups of younger and older adults were performed, controlling for gender (see Supplementary Material).

(iv) Associations with Depression, Decision-making, Social Cognition and Personality

Additional associations were explored between the AI and non-mnemonic constructs that have been related to AM abilities in past studies. Partial product-moment correlations were conducted between the eight AI measures and: depressive symptoms, temporal discounting, and measures of social cognition and personality, controlling for age and gender. Analyses were conducted in the full sample with the exception of depression, which was measured with different questionnaires in younger and older adults. Repeat analyses within each age group controlling for gender can be found in Supplementary Material.

(v) Confirmatory Factor Analysis of the Autobiographical Interview

A primary feature of declarative memory is the distinction between episodic and semantic memory (Tulving, 1972). This feature is embedded within the internal and external distinction of the AI. We tested this two-factor model of the AI with a confirmatory factor analysis (CFA), which divided all of the AI sub-categories into latent variables of internal and external details. We estimated two models across participants (with age-group embedded within the model): one with detail count and the second with detail density. Since some categories of details are reported more than others, we transformed the data by taking the cubed root of all values. CFAs were conducted with maximum likelihood estimation. Latent factors were standardized for free estimation of factor loadings. Model fit was evaluated with the following statistics: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). Finally, we examined model fit as a function of count or density values. To do so, we compared the closeness of our two models to the data's true structure using a Vuong closeness test for non-nested models (Vuong, 1989). Our hypothetical CFA model structure is depicted in Figure 1.

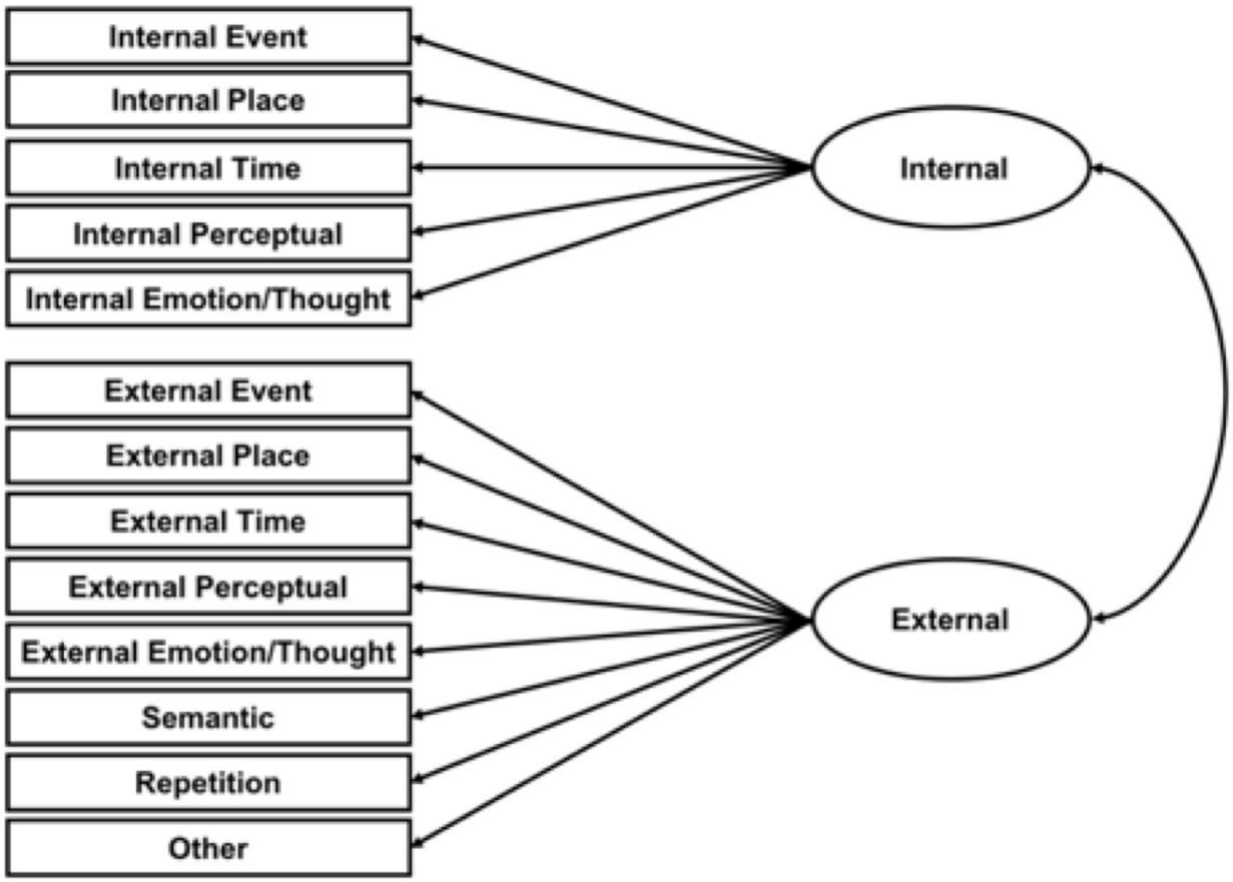



*Figure 1.* Confirmatory Factor Analysis of Internal and External Details: Hypothetical Model. The ellipses represent the latent variables of internal and external details whereas the rectangles represent observed sub-categories of detail types which fall into each latent variable category, as represented by the straight arrows. The curved arrows represent the correlation between the two latent variables.

## Results

(i) Reliability of the Autobiographical Interview

*Inter-Rater Reliability*. Based upon the rigorous standardized protocol for training scorer, we predicted high inter-rater reliability. Indeed, inter-rater reliability of the AI was high for internal and external details. Reliability of the average detail counts, as estimated using an intraclass correlation mean-rating model, was high across the entire sample (total $r(351)=.94$, $p<.001$; internal $r(351)=.95$, $p<.001$; external $r(351)=.86$, $p<.001$), within younger adults (total $r(200)=.96$, $p<.001$; internal $r(200)=.95$, $p<.001$; external $r(200)=.88$, $p<.001$), and within older adults (total $r(150)=.92$, $p<.001$; internal $r(150)=.95$, $p<.001$; external $r(150)=.83$, $p<.001$). Reliability for sub-categories was more variable in the full sample ($r's(351)=.57 – .93$, $p<.001$), within younger adults ($r's(200) =.42 – .93$, $p<.001$), and within older adults ($r's(150)=.44 – .94$, $p<.001$). Individual sub-category mean-rating model correlations are provided in Table 2.

*Internal consistency (temporal stability)*. Event remoteness can have a significant impact on memory. Because the AI samples discrete events that vary as a function of remoteness, we assessed the temporal stability of the AI in two ways. First, we examined mean differences in each of the eight AI variables between the most recent and most remote events across all participants. This informed the stability of detail recollection with temporal distance. We predicted that remote events would have fewer details overall.

Compared with remote events, recent events had higher internal count ($t(351) = 10.71$, $p <.001$, Cohen's $d =0.57$), higher internal density ($t(351) = 2.06$, $p <.05$, Cohen's $d =0.11$), higher external count ($t(351) = 6.30$, $p< .001$, Cohen's $d =0.34$) and higher semantic count ( $t(351) = 4.69$, $p <.001$, Cohen's $d =0.25$). Word count was also higher for recent versus remote events ($t(351) = 10.84$, $p <.001$, Cohen's $d =0.58$). These differences are depicted in Figure 2.

Within younger adults, we tested for mean differences on all AI variables across childhood, teenage years, and recent early adulthood memories (Figure 3). We found a stepwise increase in internal count with recency ($F(2, 400)=35.33$, $p<.001$, Cohen's $f=0.42$; Figure 3a). Remote childhood memories also had lower external count ($F(2, 400)=13.13$, $p<.001$, Cohen's $f =0.26$; Figure 3b) and semantic count ($F(2, 400)=5.20$, $p<.01$, Cohen's $f =0.16$; Figure 3c). A stepwise increase in word count for recency was also observed ($F(2, 400)=38.91$, $p<.001$, Cohen's $f =0.44$; Figure 3h). No differences were observed for density or proportion scores ($Fs < 2$, $p's > .15$; Figures 3d and 3e).

| | | All | Younger Adults | Older Adults |
|---|---|---|---|---|
| Internal and External | Total | .94*** | .96*** | .92*** |
| Internal | Total | .95*** | .95*** | .95*** |
| | Event | .93*** | .93*** | .92*** |
| | Place | .74*** | .73*** | .76*** |
| | Time | .89*** | .89*** | .88*** |
| | Perceptual | .89*** | .86*** | .94*** |
| | Emotion/Thought | .91*** | .91*** | .92*** |
| External | Total | .86*** | .88*** | .83*** |
| | Event | .60*** | .76*** | .50*** |
| | Place | .66*** | .58*** | .64*** |
| | Time | .64*** | .48*** | .71*** |
| | Perceptual | .70*** | .42*** | .78*** |
| | Emotion/Thought | .60*** | .77*** | .44*** |
| | Semantic | .85*** | .81*** | .83*** |
| | Repetition | .59*** | .58*** | .61*** |
| | Other | .57*** | .68*** | .45*** |

*Table 2.* Inter-Rater Reliability. Intraclass correlations estimating reliability of detail counts when averaged between raters using a mean-rating model. Correlations are reported for all participants, younger adults, and older adults for summary and sub-category detail counts. n.s. not significant; *p<.05; **p < .01; ***p<.001.

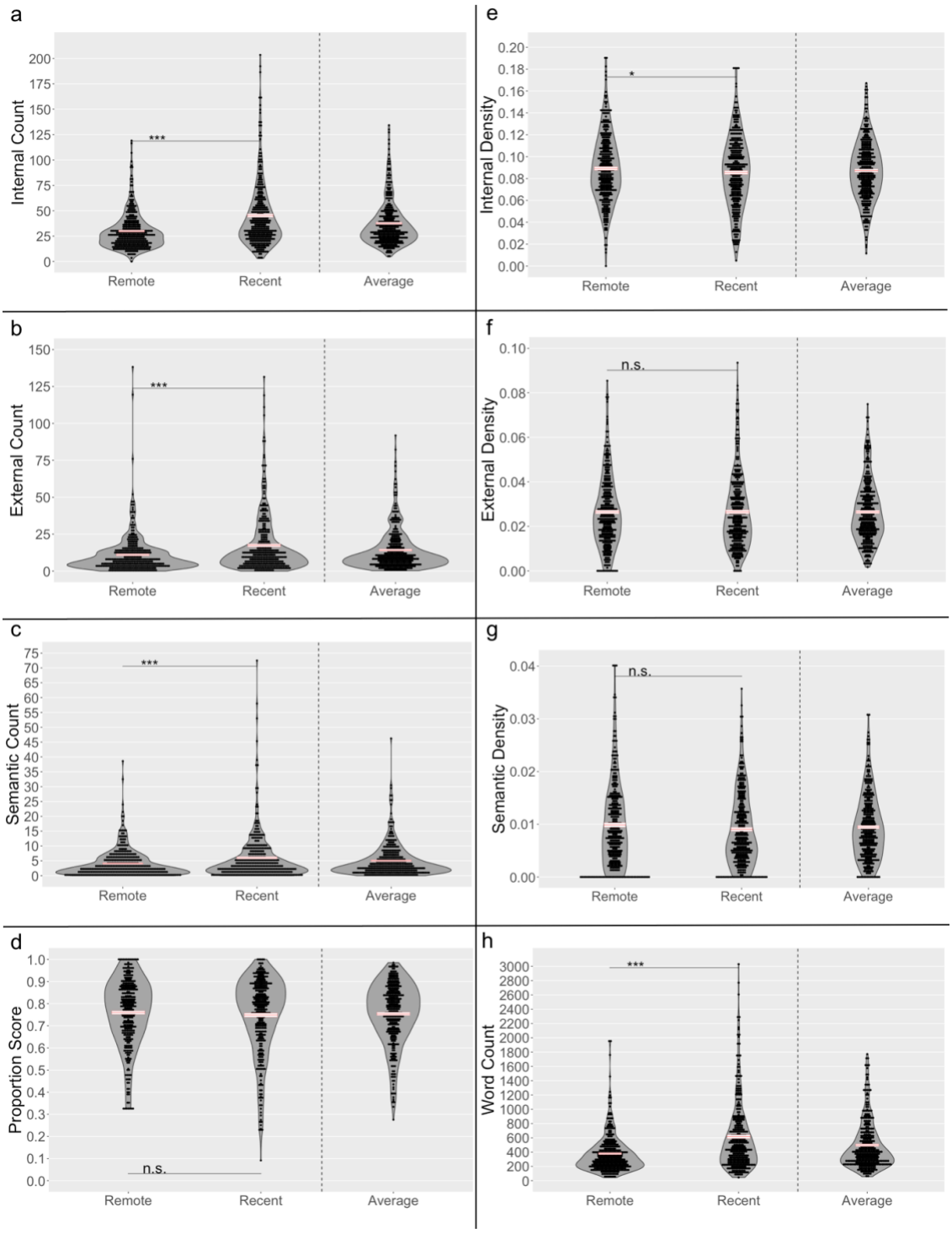



*Figure 2.* Between Memory Consistency for All Participants. The a) internal, b) external, and c) semantic counts, d) the proportion score, the density scores for e) internal f) external and g) semantic details, and h) the word count averaged across raters at the time points of most remote and most recent memories for all participants. The averages across all memories are presented for comparison but excluded from analyses. Error bars represent the mean +/- standard error. Significant differences between recent and remote memories were identified using student's paired t-test; n.s. not significant; *p<.05; **p < .01; ***p<.001.

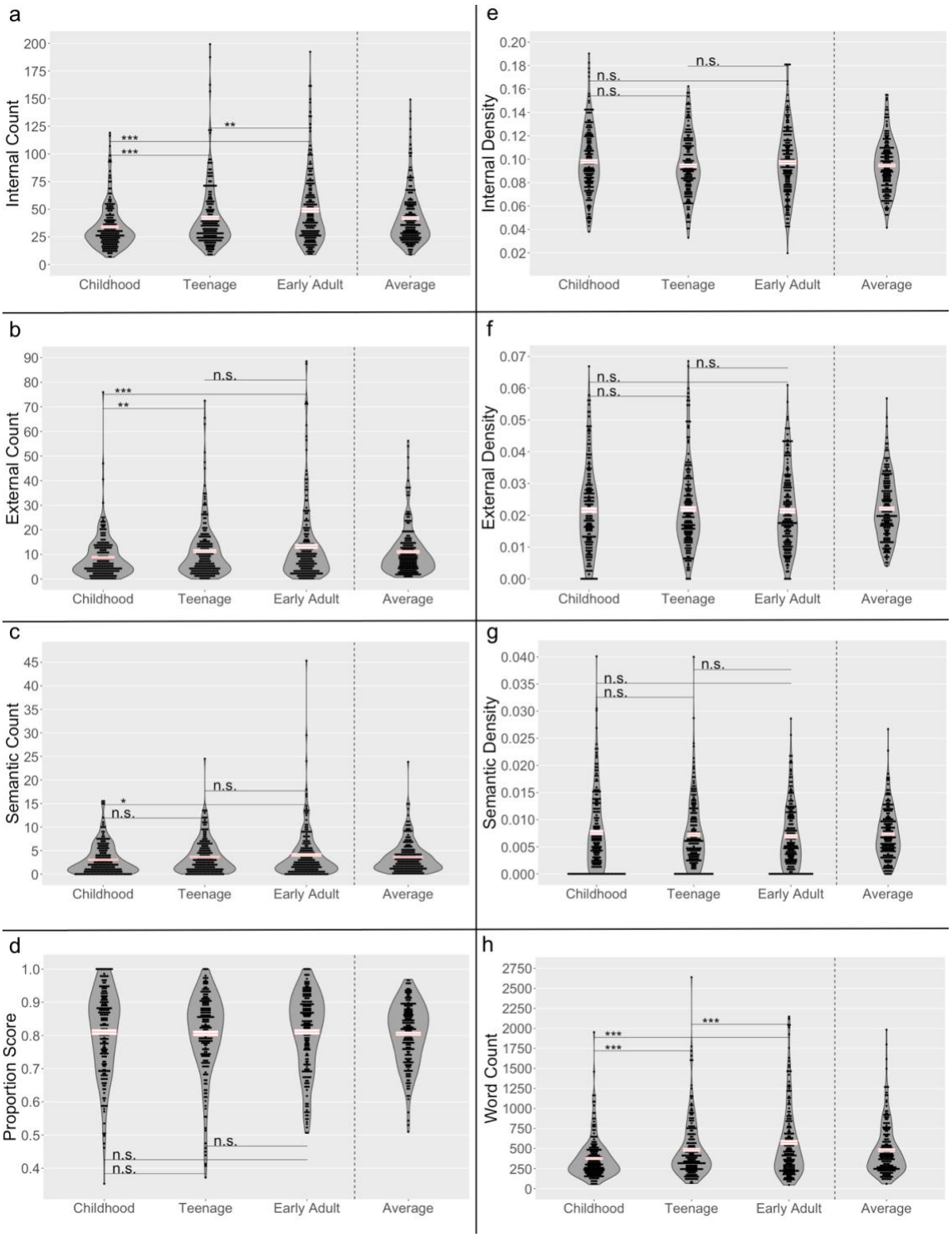



*Figure 3.* Between Memory Consistency for Younger Adults. The a) internal, b) external, and c) semantic counts, d) the proportion score, the density scores for e) internal f) external and g) semantic details, and h) the word count averaged across raters at the time points of childhood, teenage, and early adult memories for younger adults. The averages across all memories are presented for comparison but excluded from analyses. Error bars represent the mean +/- standard error. Significant differences were identified using Bonferroni corrected pairwise t-tests; n.s. not significant; *p<.05; **p < .01; ***p<.001.

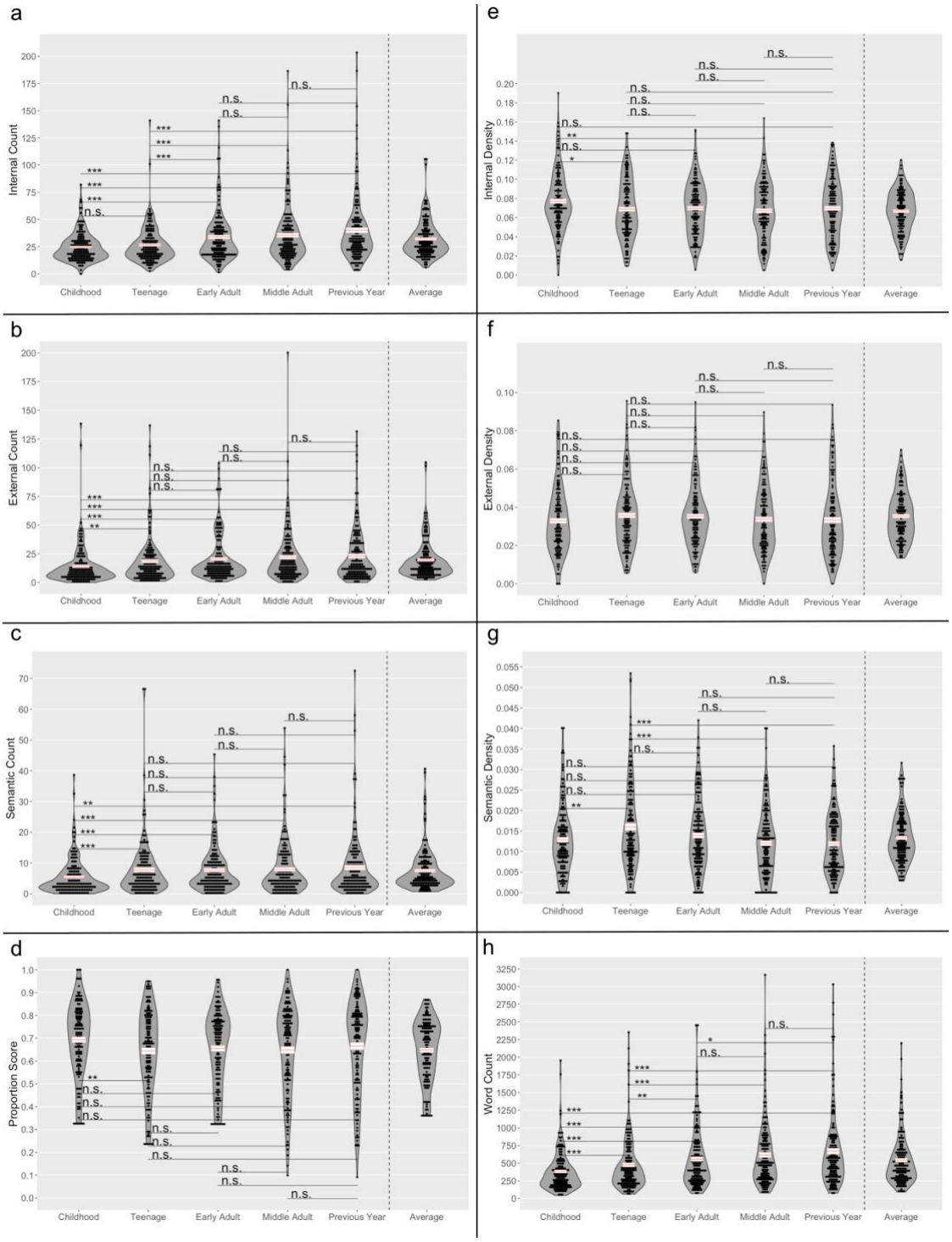



*Figure 4.* Between Memory Consistency for Older Adults. The a) internal, b) external, and c) semantic counts, d) the proportion score, the density scores for e) internal f) external and g) semantic details, and h) the word count averaged across raters at the time points of childhood, teenage, early adult, middle adult, and previous year events for older adults. The averages across all memories are presented for comparison but excluded from analyses. Error bars represent the mean +/- standard error. Significant differences were identified using Bonferroni corrected pairwise t-tests; n.s. not significant; *p<.05; **p < .01; ***p<.001.

Within older adults, we tested for mean differences on all AI variables across childhood, teenage years, early adulthood, middle age, and within the last year (Figure 4). Significant differences between the five events were found, with a pattern of lower values for remote events in internal count ($F(4,600)=20.95, p<.001$, Cohen's $f=0.37$; Figure 4a), external count ($F(4,600)=8.71, p<.001$, Cohen's $f=0.24$; Figure 4b), and semantic count ($F(4,600)=6.15, p<.001$, Cohen's $f=0.20$; Figure 4c). Lower word count was observed in more remote events ($F(4,600)=27.14, p<.001$, Cohen's $f=0.43$; Figure 4h). Differences were also observed for internal density ($F(4,600)=4.05, p<.01$, Cohen's $f=0.16$; Figure 4e) and proportion score ($F(4,600)=2.84$, $p<.05$, Cohen's $f=.14$; Figure 4e), with higher proportion and internal densities for childhood memories relative to the teenage years. Semantic density also differed ($F(4,600)=8.17, p<.001$).

Recollection on the AI was robustly correlated across time periods. Recent and remote events were highly correlated across the entire sample (r's (351) = .35 – .56, $p < .001$), in younger adults (r's (200) = .30 – .61, $p < .001$), and in older adults (r's (150) = .20 – .49, $p < .05$). Correlations between each individual event and the average score of all events (akin to an item-total correlation) were observed across the entire sample (r's (351)= .68-.84, $p < .001$), in younger adults (r's (200)= .63 – .91, $p < .001$), and in older adults (r's (150)= .54 – .90, $p < .001$). Correlations between each individual event and the average of the most recent and most remote event were also observed for the entire sample (r's (351)= .76 – .94, $p < .001$). This association was significant but highly variable for both the younger (r's (200)= .19-.95, $p < .01$) and older groups (r(150)= .35 – .96, $p < .001$). Finally, we assessed the similarity of two composite measures: the average of all events and an average of the most recent and remote events. These correlations were highly positive across the entire sample (r's (351)= .84 – .93, $p < .001$), in younger adults (r's (200)= .83 – .95, $p < .001$), and in older adults (r(150)= .77 – .91, $p<.001$). Correlations between event-level and composite AI measures are shown in Table 3.

As predicted, detail recollection varied with memory age. Importantly, the observed correlations for recollection performance over time suggest that the process of event recollection, as assessed by the AI, is consistent across all event memories, irrespective of memory age. This provides empirical support for the decision to average details over all events. The pattern of significant associations did not change when controlling for gender.

| All Participants | Internal Count | Internal Density | Proportion Score | External Count | Semantic Count | External Density | Semantic Density | Word Count |
|---|---|---|---|---|---|---|---|---|
| Remote-Recent | .55*** | .36*** | .35*** | .37*** | .38*** | .35*** | .26*** | .55*** |
| Remote-Average | .79*** | .65*** | .64*** | .78*** | .75*** | .63*** | .66*** | .80*** |
| Recent-Average | .84*** | .80*** | .76*** | .75*** | .76*** | .74*** | .62*** | .83*** |
| Remote-Remote/Recent | .80*** | .82*** | .80*** | .75*** | .72*** | .82*** | .83*** | .79*** |
| Recent-Remote/Recent | .94*** | .82*** | .84*** | .89*** | .92*** | .83*** | .76*** | .94*** |
| Average-Remote/Recent | .92*** | .88*** | .86*** | .91*** | .89*** | .84*** | .81*** | .92*** |
| **Younger Adults** | **Internal Count** | **Internal Density** | **Proportion Score** | **External Count** | **Semantic Count** | **External Density** | **Semantic Density** | **Word Count** |
| Childhood-Early Adult (Remote-Recent) | .59*** | .50*** | .37*** | .47*** | .40*** | .34*** | .30*** | .60*** |
| Childhood-Average | .82*** | .74*** | .67*** | .75*** | .70*** | .63*** | .69*** | .83*** |
| Teenage-Average | .89*** | .76*** | .70*** | .82*** | .77*** | .69*** | .64*** | .91*** |
| Early Adult-Average | .86*** | .84*** | .76*** | .87*** | .87*** | .74*** | .67*** | .88*** |
| Childhood-Remote/Recent | .83*** | .86*** | .84*** | .76*** | .74*** | .84*** | .85*** | .83*** |
| Teenage-Remote/Recent | .69*** | .49*** | .29*** | .60*** | .55*** | .24*** | .19** | .75*** |
| Early Adult-Remote/Recent | .94*** | .87*** | .82*** | .93*** | .91*** | .80*** | .76*** | .95*** |
| Average-Remote/Recent | .94*** | .92*** | .86*** | .95*** | .95*** | .83*** | .84*** | .95*** |
| **Older Adults** | **Internal Count** | **Internal Density** | **Proportion Score** | **External Count** | **Semantic Count** | **External Density** | **Semantic Density** | **Word Count** |
| Childhood-Previous Year (Remote-Recent) | .49*** | .19* | .34*** | .34*** | .38*** | .35*** | .23** | .49*** |
| Childhood-Average | .70*** | .54*** | .62*** | .79*** | .76*** | .63*** | .64*** | .78*** |
| Teenage-Average | .74*** | .66*** | .64*** | .82*** | .85*** | .62*** | .66*** | .87*** |
| Early Adult-Average | .83*** | .76*** | .71*** | .85*** | .84*** | .64*** | .69*** | .90*** |
| Middle Adult-Average | .77*** | .71*** | .69*** | .84*** | .86*** | .62*** | .73*** | .89*** |
| Previous Year-Average | .81*** | .74*** | .77*** | .69*** | .72*** | .74*** | .58*** | .78*** |
| Childhood-Remote/Recent | .73*** | .78*** | .77*** | .75*** | .71*** | .80*** | .81*** | .76*** |
| Teenage-Remote/Recent | .56*** | .55*** | .49*** | .61*** | .63*** | .41*** | .44*** | .69*** |
| Early Adult-Remote/Recent | .57*** | .54*** | .51*** | .65*** | .62*** | .40*** | .35*** | .69*** |
| Middle Adult-Remote/Recent | .49*** | .45*** | .48*** | .65*** | .62*** | .39*** | .43*** | .66*** |
| Previous Year-Remote/Recent | .96*** | .77*** | .86*** | .87*** | .92*** | .85*** | .76*** | .94*** |
| Average-Remote/Recent | .87*** | .82*** | .86*** | .89*** | .87*** | .84*** | .77*** | .89*** |

*Table 3.* Between Memory Consistency of Events and Composite Measures. Partial product-moment correlations of the AI between events and composite measures. Correlations are reported for all participants controlling for age group and gender, younger adults controlling for gender, and older adults controlling for gender. n.s. not significant; *p<.05; **p < .01; ***p<.001.

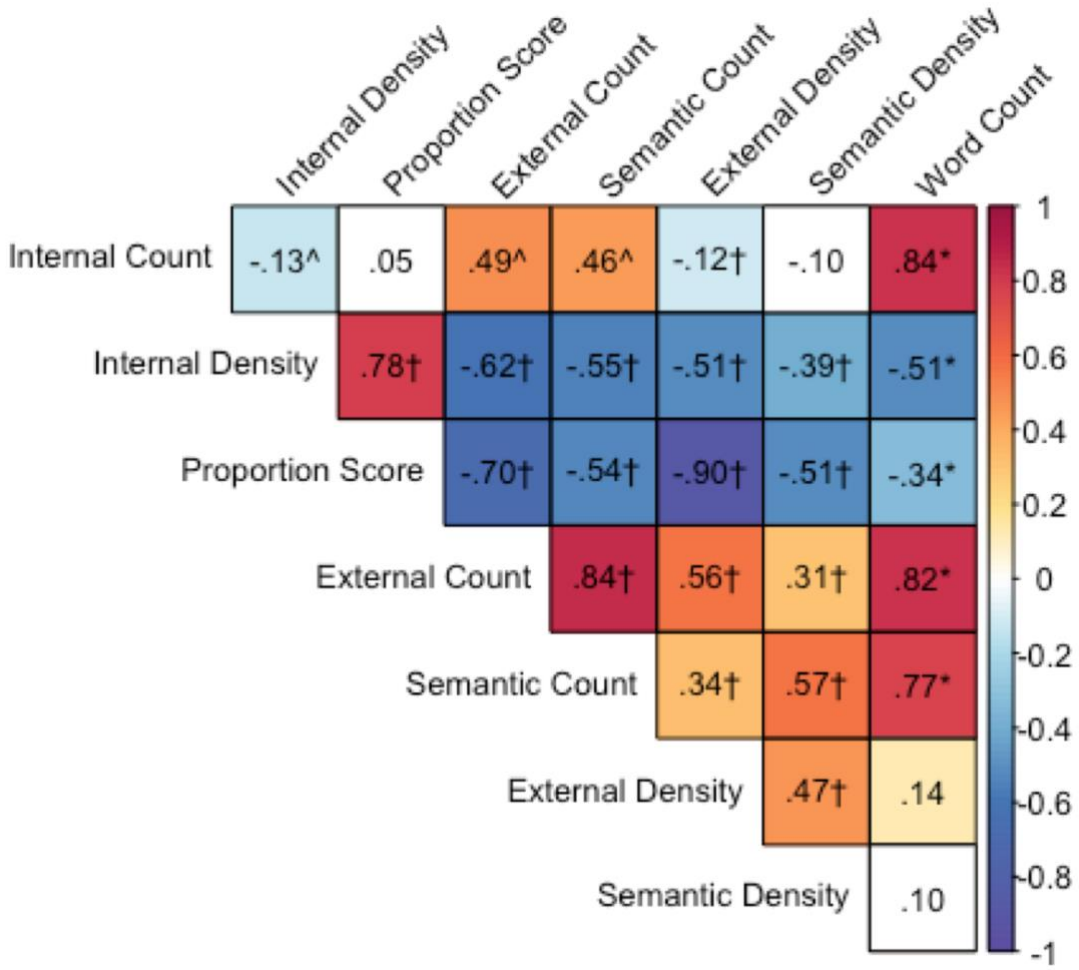



*Figure 5.* Internal Consistency of the Autobiographical Interview: Measures of Interest. Partial product-moment correlations between the 8 AI measures of interest across all participants controlling for age and gender. The matrix contains r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction, ^ associations opposite to prediction ($p < .05$, uncorrected).

*Internal consistency (AI metrics).* We next investigated associations among all AI variables. Among the primary AI metrics, we predicted that internal measures would be positively correlated, and that external and semantic measures would be correlated (although this is largely mandated, as semantic details are included in external scores). We predicted that internal and external counts would be correlated with word count, but that word count would not be associated with proportion or density scores.

When examining correlations among AI variables three primary patterns emerged (see Figure 5 for full results). First, all detail count measures were positively correlated. Second, internal density and proportion scores, which adjust for verbosity, were positively correlated, and both were negatively correlated with the external detail measures (count and density). Third, external measures were positively correlated with each other and word count. All patterns held within age groups (Supplemental Figure 1). Contrary to predictions, internal count and internal density were negatively correlated across participants (pr(348)= -.13, p< .05) and within younger adults (pr(198)= -.21, p< .01), but were uncorrelated within older adults (pr(148)= .02, p= .77).

Our observation of a positive correlation between internal and external counts is inconsistent with a recent study reporting that these measures were negatively correlated (Devitt et al., 2017). To explore this association further, we conducted a hierarchical linear model analysis (see Methods). The regression weight for internal details was significant ($B$ = 0.52, $SE$=0.05, $t$ ratio = 11.24, $df$ =1130, $p$ < .001), indicating that events reported with more internal details were more likely to include more external details. The age group regression weight was also significant ($B$ = 2.12, $SE$=0.34, $t$ ratio = 6.21, $df$=1160, $p$ < .001), indicating that older adults were more likely to provide more external details, as we and others have shown in earlier reports (Levine et al., 2002; Spreng et al., 2018). The interaction term was also significant ($B$ = -0.20, $SE$=0.06, $t$ ratio =-3.21, $df$=1271, $p$ < .01). Internal details were less predictive of external details for older versus younger adults. These analyses provide additional evidence that internal count is positively associated with external (as well as semantic) detail counts. This association remains significant when accounting for age group and within-participant variance.

*Internal consistency (AI ratings).* The full AI protocol includes self-reported participant ratings and scorer-assigned ratings for level of detail. These self-report and summary rating scores complement the primary focus on the core AI metrics.

Self-report ratings include event vividness, emotional change, significance of the event now and when it originally occurred, and rehearsal frequency. We predicted that internal scores would positively correlate with self-report ratings (event vividness, emotional change, significance now, significance then, rehearsal frequency). All ratings were positively correlated with internal count, external count, and total word count (Figure 6). Contrary to predictions, these ratings, with the exception of rehearsal, were negatively correlated with internal density and the proportion measures. No associations were

observed with rehearsal that controlled for verbosity (density, proportion). Similar patterns were observed within age groups (Supplemental Figure 2).

In addition, scorers are asked to assign an overall score for each memory based on details for place, time, perceptual, emotion/thought, AMI, episodic richness (internal) and time integration (external). External details are only considered for the time integration rating. We examined associations between the AI metrics and the overall ratings assigned by scorers. Our predictions follow the scoring guide, with internal scores predicted to be associated positively with place, time, perceptual, emotion/thought, AMI and episodic richness, while external scores will positively correlate with the time integration rating. Generally, scorer ratings were positively correlated with all counts while associations with density and proportion measures were mixed. Internal measures controlling for verbosity did not show many positive associations with ratings tracking internal detail. Unlike with counts, external density was the only density measure positively correlated with time integration. See Figure 7 for full results. We found similar results within age groups (Supplemental Figure 3) with one notable exception: in younger adults, AMI rating was negatively correlated with internal density (partial $\rho(198)= -.18$, $p< .01$) and proportion score (partial $\rho(198)= -.14$, $p< .05$), and episodic richness was negatively correlated with proportion score (partial $\rho(198)= -.18$, $p< .05$). In older adults, AMI rating was positively correlated with internal density (partial $\rho(148)= .17$, $p< .05$) and proportion score (partial $\rho(148)= .42$, $p< .001$), and episodic richness was also positively correlated with the proportion score (partial $\rho(148)= .32$, $p< .001$).

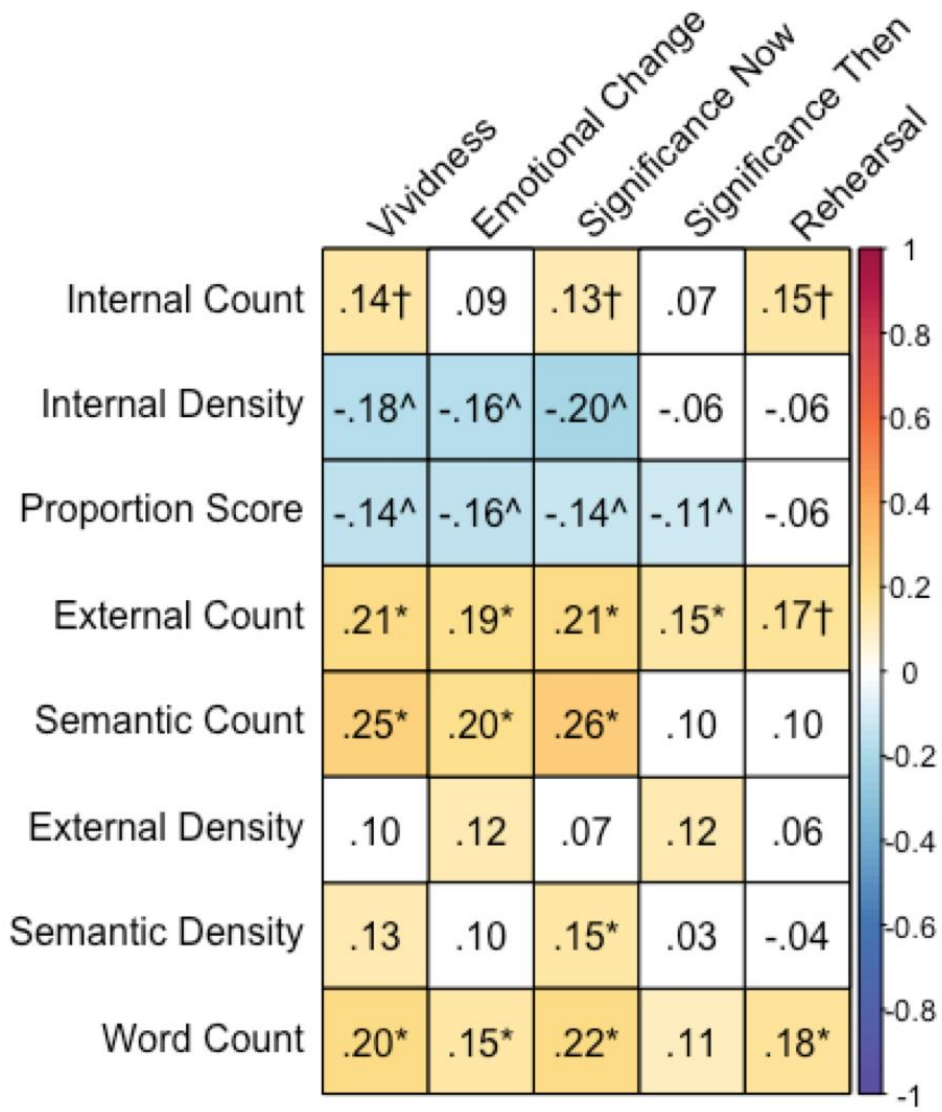



*Figure 6.* Internal Consistency of the Autobiographical Interview: Participant Self-Report Ratings. Partial spearman correlations between the 8 AI measures of interest and participant self-report ratings across all participants controlling for age and gender. The matrix contains partial ρ values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction, ^ associations opposite to prediction ($p < .05$, uncorrected).

(ii) Demographic Associations with the Autobiographical Interview

We next assessed how AI scores varied by demographics, including age, gender and education (Figure 8). We predicted a robust age effect with higher internal details in younger adults, and higher external details for older adults. Few reports have examined sex and gender differences in AM, and the findings to date have been mixed (e.g., Compere et al., 2016; Fuentes and Desrocher, 2013). Similarly, little work has directly assessed the impact of education on AI performance or AM more broadly. As such, we offer no specific predictions with respect to sex/gender or education differences.

The predicted main effects of age were supported, with significant age effects observed on all AI measures except for word count. Compared with older adults, younger adults had higher internal count ($F(1, 348)=18.11$, $p<.001$, Cohen's $d=0.45$), higher internal density ($F(1, 348)=144.09$, $p<.001$, Cohen's $d=1.29$), and a higher proportion of internal to total details ($F(1, 348)=172.23$, $p<.001$, Cohen's $d=1.42$). Compared to younger adults, older adults had higher external count ($F(1, 348)=35.73$, $p<.001$, Cohen's $d=0.65$), higher semantic count ($F(1, 348)=48.19$, $p<.001$, Cohen's $d=0.75$), higher external density ($F(1, 348)=126.60$, $p<.001$, Cohen's $d=1.21$), and higher semantic density ($F(1, 348)=115.94$, $p<.001$, Cohen's $d=1.16$).

We found evidence for a small gender effect ($F(1, 348)=7.47$, $p<.01$, Cohen's $d = 0.25$), where women had higher internal density than men, although men had a higher overall word count ($F(1, 348)=4.81$, $p<.05$, Cohen's $d = 0.23$). There were no significant interaction effects ($F(1, 348)=.00028$-$1.61$, $p>.21$).

There is little consensus on the relationship between education and AM with both null (Berna et al., 2012; Murre et al., 2014) and positive associations previously observed (Borrini et al., 1989; Wessel et al., 2014). A modest negative correlation was observed in the current sample between external density and years of education across all individuals ($pr(348)= -.12$, $p< .05$), but did not remain significant within either the younger or older adult groups (younger $pr(198)= -.13$, $p=.07$; older $pr(150)=-.10$, $p=.22$). No other significant associations were observed between AI scores and years of education across all participants or within age groups. Based on the small *r* value with external density and the lack of any other associations, we did not control for education in subsequent analyses (Cohen, 1988; Gignac & Szodorai, 2016).

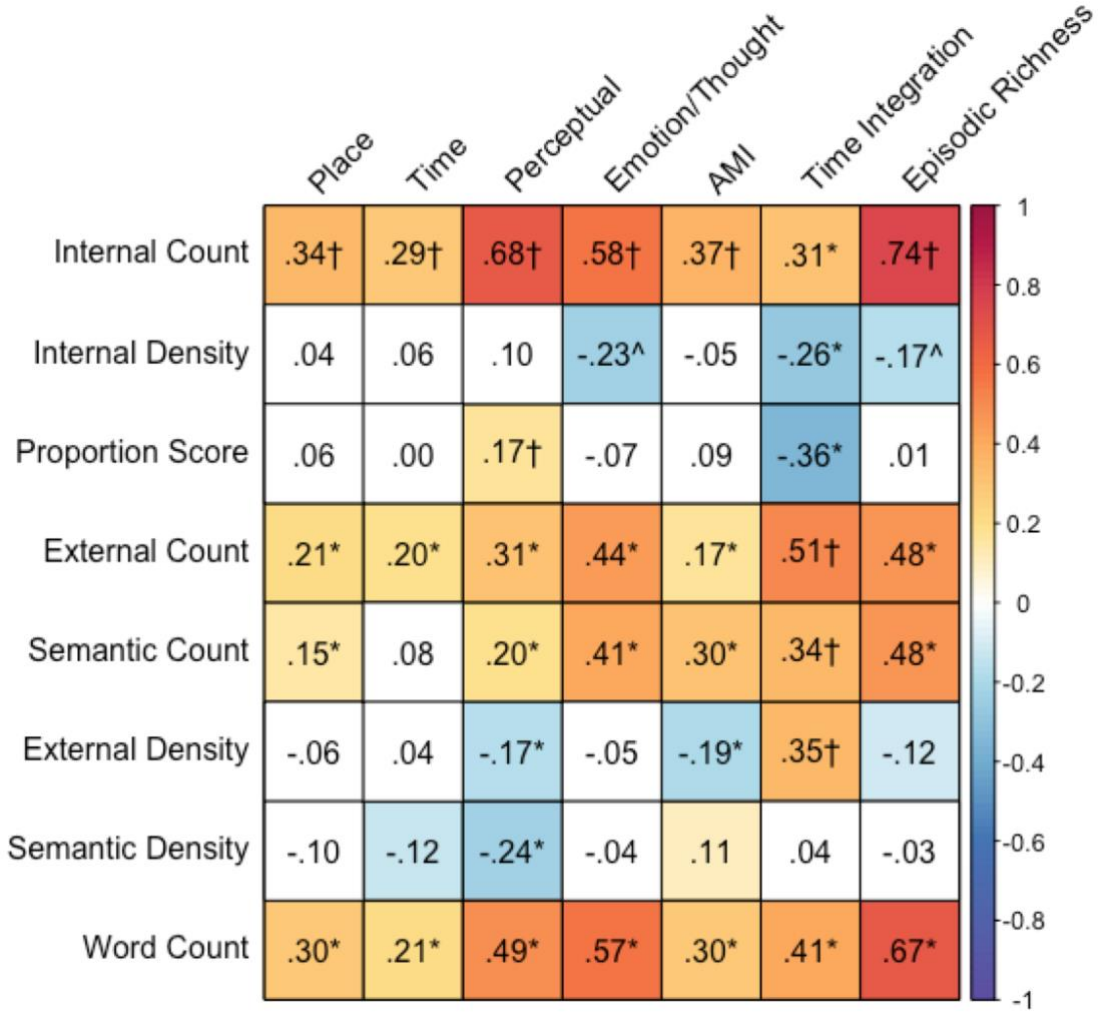



*Figure 7.* Internal Consistency of the Autobiographical Interview: Scorer Ratings. Partial spearman correlations between the 8 AI measures of interest and average scorer ratings across all participants controlling for age and gender. AMI= Autobiographical Memory Interview. The matrix contains partial ρ values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction, ^ associations opposite to prediction ($p < .05$, uncorrected).

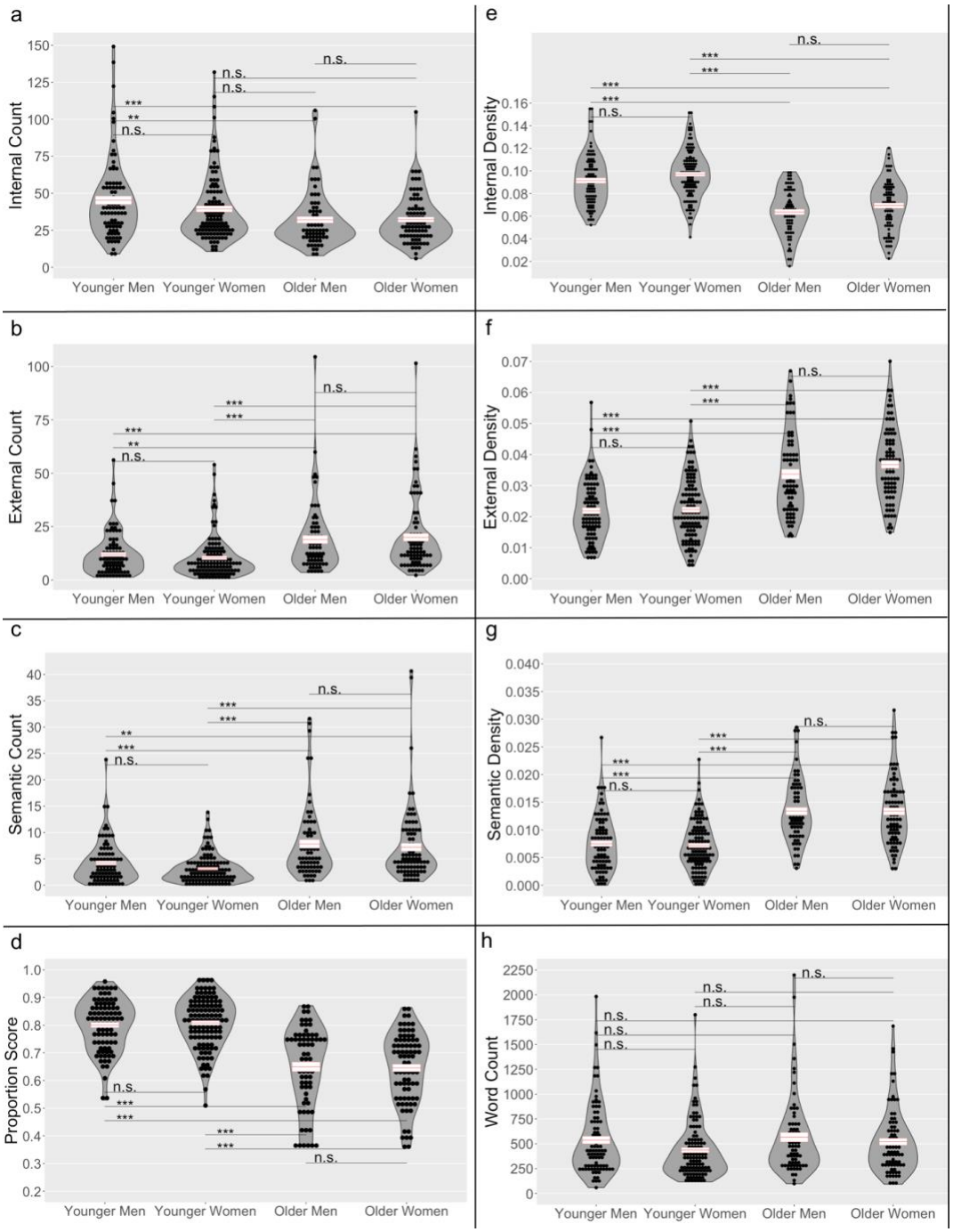



*Figure 8.* Gender and Age Differences. The a) internal, b) external, and c) semantic counts, d) the proportion score, the density scores for e) internal f) external and g) semantic details, and h) the word count averaged across raters and time periods and separated by gender and age groups. Error bars represent the mean +/- standard error. Significant differences were identified using pairwise t-tests. P values are Bonferroni corrected for tests without a-priori hypotheses; n.s. not significant; *p<.05; **p < .01; ***p<.001.

(iii) Convergent Validity of the Autobiographical Interview

We next examined the convergent validity of the AI scores with index scores of episodic memory, semantic memory, executive function and the retrieval processes of recognition and familiarity.

Consistent with the established theories of discrete memory systems, and the original framing of the AI as an instrument capable of differentiating episodic and semantic recollection, we predicted that episodic memory measures would be positively correlated with the internal detail metrics on the AI and semantic measures would positively correlate with external details. We also predicted that executive function measures would be positively correlated with internal details and negatively correlated with external details, consistent with previous work highlighting the role of executive function in the search and retrieval of episodic memories (Abellán-Martínez et al., 2019; Conway & Pleydell-Pearce, 2000; Yubero et al., 2011).

We first confirmed that cognitive functions followed expected age patterns (Park et al., 2001). Episodic index scores were significantly higher for young (Young: M=0.52, SD=0.46; Old: M=-0.70, SD=0.71; t(241) = 18.38, p < .001, Cohen's d = 2.10). Semantic index scores were significantly higher for older adults (Young: M=-0.37, SD=0.75; Old: M=0.50, SD=0.71; t(331) = 10.99, p < .001, Cohen's d =1.18). Executive function index scores were significantly higher for young (Young: M=0.35, SD=0.52; Old: M=-0.46, SD=0.51; t(326) = 14.66, p < .001, Cohen's d =1.57).

Internal AI measures were positively correlated with standard measures of episodic memory. Across the whole sample, the episodic memory index score was positively correlated with internal count (pr(348)= .24, p< .001), internal density (pr(348)= .15, p< .01), and proportion score (pr(348)= .28, p< .001). The episodic memory index score was negatively correlated with external density (pr(348)= -.25, p< .001). The semantic memory index score was not significantly associated with any of the AI measures (pr(341) range= -.06 – .05, *p*'s= .25 – .93). Consistent with predictions, the executive function index score was significantly associated with internal density (pr(348) = .10, p = .05). No other associations were observed with executive function. Correlation between the AI and index scores are shown in Figure 9. Association patterns are also evident in the individual measures that comprise the index scores as visualized in Supplemental Figure 4.

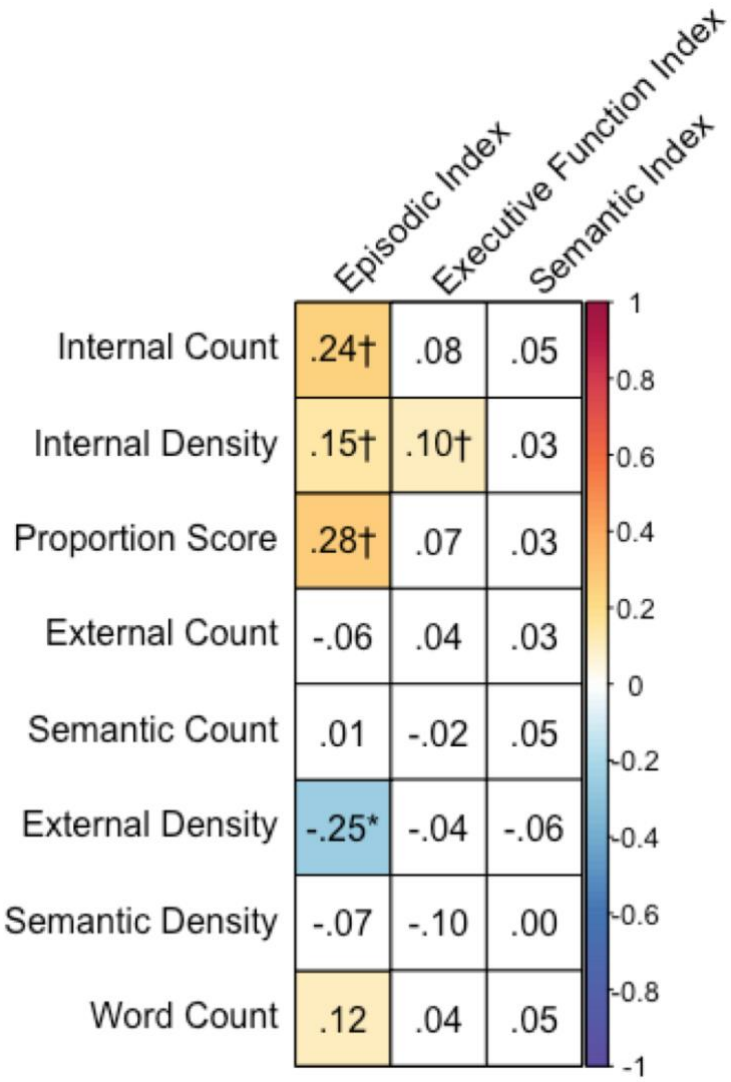



*Figure 9.* Convergent Validity of the Autobiographical Interview with Episodic, Fluid, and Semantic Index Scores. Partial product-moment correlations between the 8 AI measures of interest and the 3 episodic, fluid, and semantic Index scores across all participants controlling for age and gender. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction.

We found similar results within age groups, with some exceptions (Supplemental Figure 5). In younger adults, internal density was not associated with either the episodic index (pr(198)= .01, *p*= .90) or the executive function index (pr(198)= .08, *p*= .25) . In older adults, the executive function index was not associated with internal density (pr(148)= .13, *p*= .10) but was negatively associated with semantic density (pr(148)= -.16, *p*<.05). See Supplemental Figure 6 for association patterns between individual measures that comprise the index scores in the younger and older groups.

Drawing from previous work showing that memory specificity is positively associated with recollection and negatively associated with familiarity (Piolino et al., 2006), we predicted that internal detail measures would be positively and specifically correlated with recollection, as they putatively assess episodic re-experiencing. Contrary to our predictions, no associations with recollection were observed for the full sample (Figure 10). Negative associations were observed between familiarity and external count (pr(234)= -.13, *p*< .05), semantic count (pr(234)= -.20, *p*< .01), and semantic density (pr(234)= -.18, *p*< .01). In contrast, proportion score was positively correlated with familiarity (pr(234) range= .18, *p*< .01). No other associations were significant across the full sample. No significant correlations emerged within the younger adult group. For older adults we observed an additional positive correlation between familiarity and internal density (pr(87)= .27, *p*< .01).

(iv) Associations with Depression, Decision-Making, Social Cognition and Personality

Based on previous reports implicating non-mnemonic factors in AI performance, we examined associations between AI and measures of depression, decision-making (temporal discounting), social cognition, and personality.

*Depression.* We predicted that internal measures would be negatively correlated with measures of depression, given previous reports of an inverse relationship between low mood and memory performance (Brittlebank et al., 1993; Hitchcock et al., 2014; Kuyken & Dalgleish, 1995; Liu et al., 2013; Williams & Broadbent, 1986; Williams & Scott, 1988; Williams et al., 2007; Wilson & Gregory, 2018). Because our mood measurements differed for younger and older adults (BDI, GDS), we analyzed these associations in each group separately.

In younger adults, depression scores were negatively associated with proportion score (pr(115) = -.19, p < .05) and positively associated with external count (pr(115) = .22, p < .05) and external density (pr(115) = .18, p < .05). No significant associations were found between the AI and depression scores in older adults. Correlations between the AI and measures of depression are shown in Figure 11.

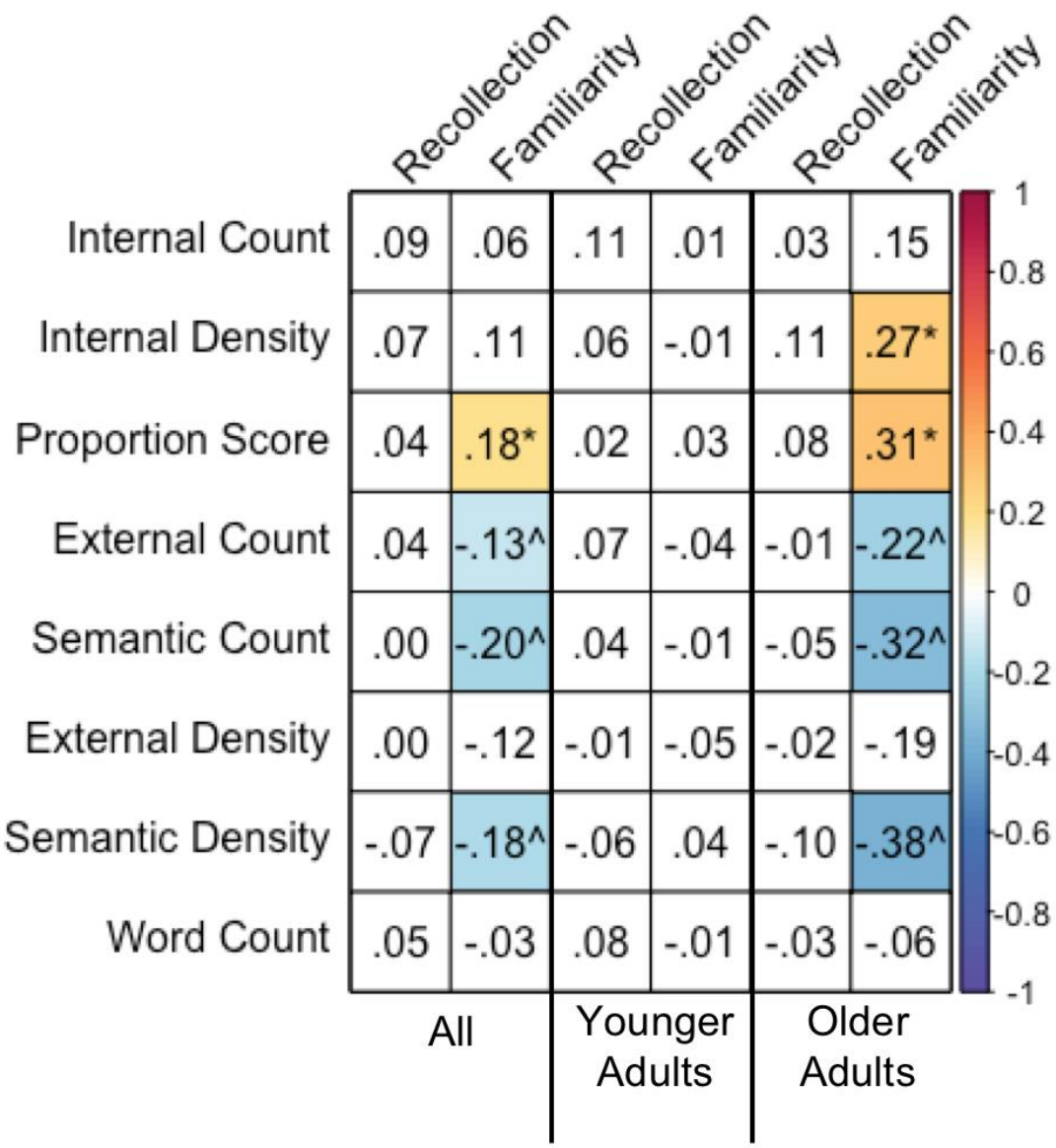



*Figure 10.* Convergent Validity of the Autobiographical Interview with Measures of Recollection and Familiarity. Partial product-moment correlations between the 8 AI measures of interest and measures of recollection and familiarity from the Remember-Know paradigm across all participants controlling for age and gender, in younger adults controlling for gender, and in older adults controlling for gender. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction, ^ associations opposite to prediction ($p < .05$, uncorrected).

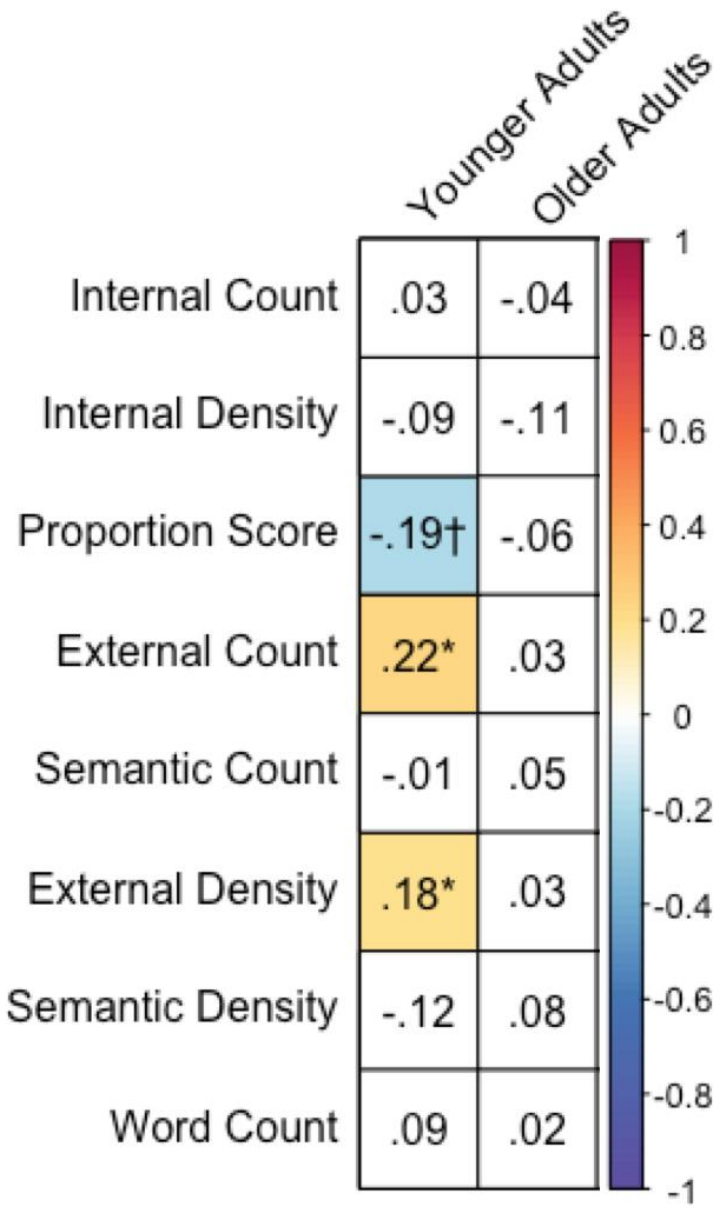



*Figure 11.* Convergent Validity of the Autobiographical Interview with Measures of Depression. Partial product-moment correlations between the 8 AI measures of interest and the Beck Depression Inventory in younger adults and the Geriatric Depression Scale in older adults across participants controlling for gender. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction.

*Decision-making (temporal discounting)*. We predicted that internal detail measures would be negatively correlated with temporal discounting based on recent work by Lempert and colleagues (2020), which suggested that the propensity to wait for larger rewards is related to episodic AM capacity. We did not observe an association between temporal discounting and AI measures across participants or within age groups (Supplemental Figure 7). Temporal discounting was not associated with the ratio of perceptual/gist-based recollection reported in Lempert et al. (2020).

*Social Cognition*. For measures of social cognition, we predicted positive correlations with internal detail measures given previously reported associations between theory of mind, sociality and AM (Buckner & Carroll, 2007; Buckner et al., 2008; Gaesser, 2013; Gaesser & Schacter, 2014; DuPre et al., 2016; Rabin et al., 2010; Spreng et al., 2009; Spreng & Grady 2010; Spreng & Mar, 2012). Consistent with predictions, across all participants, performance on the Reading of the Mind in the Eyes task was positively associated with proportion score (pr(300)= .16, $p < .01$) and negatively correlated with external density scores (pr(300)= -.21, $p < .01$; Figure 12). A similar pattern was observed in younger adults (Supplemental Figure 8a). For older adults, associations with a second measure of social cognition emerged. Perspective taking on the Interpersonal Reactivity Index was positively correlated with internal count (pr(132)= .19, $p < .05$) and semantic count (pr(132)= .24, $p < .01$; Supplemental Figure 8b).

*Personality*. We predicted that internal detail measures would be positively correlated with trait openness/intellect based on previous work linking trait openness with autobiographical coherence, narrative complexity, vividness, sensory experience, and reliving (Adler et al., 2007; McAdams et al., 2004; Rasmussen & Berntsen, 2010; Rubin & Siegler, 2004). No associations survived Bonferroni correction across participants (Figure 13) or in younger adults (Supplemental Figure 9a). In older adults (Supplemental Figure 9b), openness/intellect was positively associated with internal count (pr(133)= .18, $p< .05$) and proportion score (pr(133)= .17, $p< .05$).

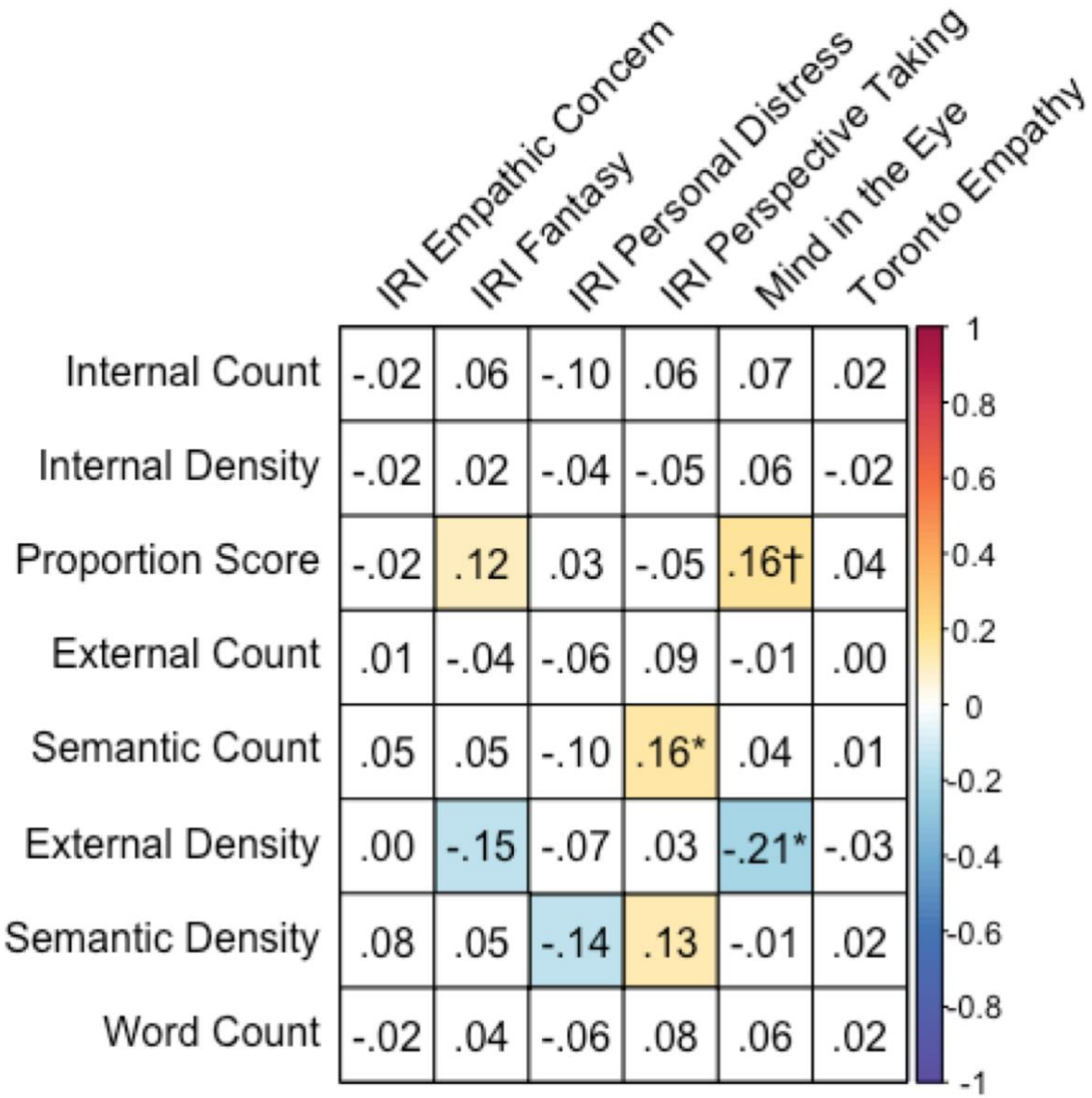



*Figure 12.* Associations between the Autobiographical Interview and Measures of Social Cognition. Partial product-moment correlations between the 8 AI measures of interest and measures of social cognition across all participants controlling for age and gender. The matrix contains partial r values, which are color-coded by strength of the correlation. The matrix contains r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction.

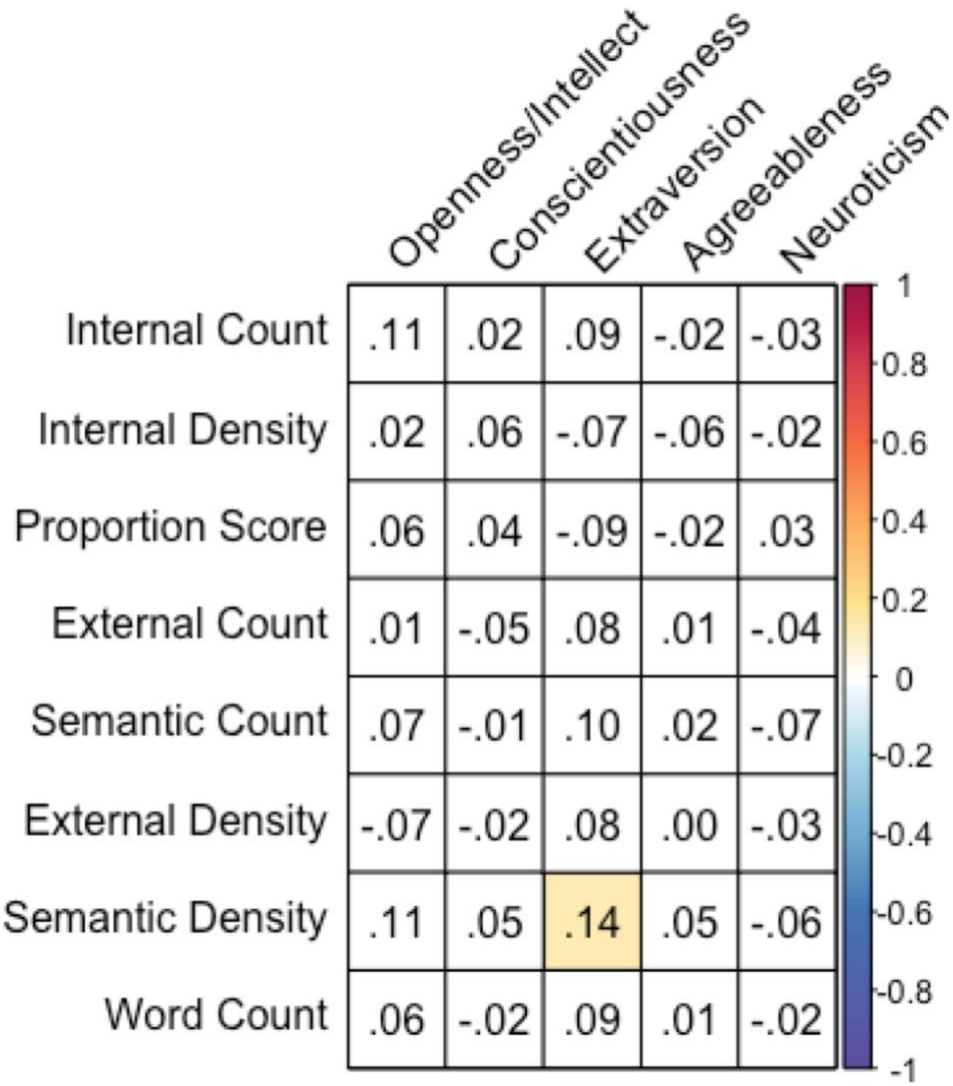



*Figure 13.* Associations between the Autobiographical Interview and Measures of Personality. Partial product-moment correlations between the 8 AI measures of interest and measures of personality across all participants controlling for age and gender. The matrix contains partial r values. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$).

(v) Confirmatory Factor Analysis

Based upon the intended design of the AI to dissociate episodic from semantic features of AM, we predicted that a two-factor model of the AI would be supported by the data. However, a two-factor model for internal and external details was not supported in the CFA. Low model fit was observed for both count and density metrics.

The CFA with counts showed a low model fit: CFI=.76, TLI= .71, RMSEA= .155 (90% CI = [.143, .167]). Recommended thresholds for good model fit are CFI at or above .95 to surpass performance of a null model (Hooper et al., 2008), TLI at or above .95 (Hooper et al., 2008), and RMSEA at or below .08 (Browne & Cudeck, 1993). This RMSEA value indicates a large difference between the residuals of the sample covariance matrix and the hypothesized covariance model. Observed variables all showed significant positive factor loadings (see Supplemental Figure 10). The two latent variables (internal/external) were also positively correlated in younger adults ($r(199)=.78$, $p< .001$) and in older adults ($r(149)= .37$, $p< .001$).

The CFA with density scores also showed a low model fit: CFI= .64, TLI= .56, RMSEA= .130 (90% CI = [.118, .142]). See Supplemental Figure 11 for factor loadings. The two latent variables (internal/external) were negatively correlated in younger ($r(199)= -.29$, $p< .01$) and in older adults ($r(149)= -.50$, $p< .001$), broadly consistent with a dissociation between internal and external features of AM. A Vuong closeness test suggested that the CFA model with density scores was closer to the true data generating process than the CFA with counts ($z = 154.18$, $p < .001$).

## Discussion

The Autobiographical Interview is a performance-based measure of autobiographical memory that has emerged as the gold standard for quantifying and characterizing personal recollective experiences. Despite the prominence of the AI, there has yet to be a comprehensive evaluation of its psychometric properties. Here we report the findings of a psychometric assessment of the AI in a well-powered sample of healthy younger and older adults. AI performance is reported using both the original approach (detail counts and proportion scores; Levine et al., 2002) and novel metrics wherein detail counts are scaled by verbosity (density scores). Our first aim was to determine whether the AI was a reliable and valid measure of AM. Inter-rater reliability and internal consistency suggested high reliability, but this depended on the metric used. Density scores were more stable across recalled events and demonstrated the expected relationship between internal and external details. Reliability was further confirmed with robust age effects across detail counts, proportion scores, and density scores. Other demographic factors had more modest effects and appear to be less replicable across samples. Convergent validity analyses provided strong evidence that the AI is a valid measure of episodic, but not semantic AM. Performance on standard episodic memory tasks was correlated with all internal detail metrics. No associations were observed with standard semantic memory measures. Density scores also correlated with executive function measures, suggesting added sensitivity of verbosity-controlled metrics to AM processes. Given the rich literature on non-mnemonic influences on AI performance, our second aim was to replicate previously reported associations between the AI and individual difference measures of psychological and cognitive function. Associations between depression and the AI were observed in the younger adults. Few other predicted associations with these non-mnemonic factors were observed. An important theoretical assumption underlying the design of the AI is that separate scoring of internal and external details disambiguates episodic from non-episodic event content. Our third aim was to empirically test this assumption. The two-factor model was not strongly supported by our data, but model fit was improved with density scores. As each of the key analyses in our assessment was performed on original and novel AI metrics, our final aim was to evaluate the merits of controlling for verbosity. Overall, we find that controlling for verbal output reveals a more stable pattern of AM performance, increased convergent validity with standard laboratory tasks of cognitive functioning, and improved correspondence to the proposed two-factor model of AM. We conclude by discussing the importance of considering verbosity when interpreting performance on the AI.

### Autobiographical Interview is highly reliable

*Inter-rater reliability*. A significant challenge for quantifying AM is the objective evaluation of another's memory. Determining what comprises a recollected event and the identification of information units within an event narrative is subjective. However, our findings demonstrate that the AI's rigorous administration and scoring protocol are highly effective at reliably parsing recollective experiences into internal and external details. This is critically necessary for characterizing and quantifying the episodicity of a recollected experience (Levine et al., 2002). Inter-rater reliability on the AI scoring was

high, far exceeding established benchmarks in the field (e.g., .75-.80; Cicchetti, 1994; Orwin, 1994). The inter-rater reliability we observe is comparable to previously reported intraclass correlation values in smaller samples which range from .88-.96 for internal details and .79-.96 for external details (Addis et al., 2008; Addis et al., 2010; Cole et al., 2012; Devitt & Schacter, 2020; Gaesser et al., 2011; Levine et al., 2002; Robin & Moscovitch, 2017; Terrett et al., 2016). While not the focus of the current report, inter-rater reliability was similarly high for internal detail sub-categories as well as for semantic details. These findings suggest that core features of autobiographical re-experiencing can be reliably quantified using the AI.

*Internal consistency (temporal stability)*. We observed high internal consistency across events recalled from different time periods. While AM clearly varies as a function of time (e.g. Conway & Pleydell-Pearce, 2000; Crovitz & Schiffman, 1974; Schroots et al., 2004), our findings suggest that stable individual differences exist in AM capacity that are observable across all time periods. As such, averaging measures across events may be an effective strategy for individual differences research involving the AI. Many studies that have utilized the AI compute averages of detail counts across memories, which assumes temporal stability in detail generation. However, this assumption is rarely tested. Our findings identified a robust recency effect for count metrics. In contrast, verbosity corrected measures did not increase with event recency (although some differences were observed in older adults). As such, verbosity corrected measures (e.g., density and proportion) may be more appropriate metrics for trait-level memory capacity, as we discuss in greater detail below. Recent and remote events were also correlated for each AI measure across the entire sample and within each age group. Robust positive associations were observed between individual events and the two composite scores (i.e., average of most recent and most remote; average of all events). This provides strong evidence that each event is measuring a similar latent construct. This finding also suggests that one event may offer sufficient estimation of AM capacity. If validated, this would broaden the feasibility of implementing the AI across a larger range of studies where time and/or sample sizes have precluded its usage.

Internal consistency

A core motivation for the development of the AI was to move beyond one-factor self-report scales of AM capacity such as the Autobiographical Memory Interview (Kopelman et al., 1989, 1990). In contrast, the AI was developed to quantify and dissociate episodic memory (internal details) from other components contributing to AM (external details), including semantic memory (a sub-category of external details). Quantifying this distinction is fundamentally necessary to isolate and thus accurately characterize the episodicity of autobiographical event recall. Here we investigated whether performance metrics on the AI (detail count, detail density and proportion scores) reflect these two dimensions of AM, thereby facilitating more precise characterizations of autobiographical re-experiencing. Count measures, including the number of internal and external details, were positively correlated and, unsurprisingly, were also highly correlated with the number of words spoken. These associations were confirmed by our hierarchical linear modeling analysis. In contrast, detail density and proportion scores

revealed a different pattern of associations. Internal density was positively correlated with proportion score, and both were negatively correlated with all external measures, as would be predicted from a two-factor account (internal/external) of AM. We suggest that total words spoken may underlie the association between internal and external detail count. However, our finding of a positive association between internal and external count is inconsistent with previous works reporting no association (Addis et al., 2008; Addis et al., 2009; Addis et al., 2010; Cole et al., 2012; Gaesser et al., 2011) or a negative association (Devitt et al., 2017). We speculate that this difference may arise from underpowered samples in previous work or variation in AI administration (e.g., time limits for participant responses). The vast majority of AI studies to date have relied on internal count measures (e.g., Palombo et al., 2018; Barry et al., 2020; Armson et al., 2021 for recent examples), but our findings provide evidence that density and proportion scores may provide a better estimate of autobiographical episodicity. The modest negative correlation between internal density and internal count observed here suggests they are measuring different constructs and should not be used interchangeably.

Full scoring of the AI includes summary ratings of the AI narratives, although these are often omitted from published reports. Scorer ratings include an appraisal of the level of internal place, time, perceptual, and emotion/thought detail provided. Additional summary ratings include the AMI, which denotes the specificity of the event chosen, time integration, which conveys the amount of global context provided for the event, and episodic richness, which reflects the impression of participant re-experiencing conveyed by the total amount of internal detail provided.

The AI scoring manual directs scorers to base ratings on internal counts, with the exception of time integration. For this reason, we predicted positive correlations would be observed between these summary ratings and internal count. Consistent with the initial validation of the AI (Levine et al., 2002), internal count was positively associated with AMI. However, most summary ratings were also positively associated with external and semantic count as well as words spoken, with mixed associations observed for verbosity-corrected measures. Time integration was correlated with external density, supporting the use of external density as a measure of non-episodic contextual details. Across the full sample, associations between scorer ratings and AI metrics were non-specific, failing to distinguish between episodic (internal) and non-episodic (external) content. As such, we caution against using scorer ratings in the place of recollection-derived internal and external measures.

Participants also rated the quality of their recollections for vividness, emotional change, significance (now and then), and rehearsal. We found that these self-report AI measures were positively associated with all count measures (internal, external, semantic, and word count), as well as semantic density. However, we interpret these associations cautiously as these may reflect non-specific associations with overall verbal output, reflecting that participants generally speak more about events that they deem to be vivid, emotional, significant, and rehearsed.

In contrast, internal density and proportion scores were negatively correlated with self-reported vividness, emotional change and significance now, mirroring overall positive correlations with external measures. The robust associations between self-report and external details may be evidence for a semantic scaffolding, conceptual integration, or narrative style: higher levels of vividness, emotional change, significance, and rehearsal could be associated with a deeper or more readily available integration into a life-story and self-concept used to access a specific event (Blagov & Singer, 2004; Wood & Conway, 2006). Indeed, these associations were stronger in older adults alone, and present only with counts in younger adults. In a similar vein, Alzheimer's disease patients, who provide fewer episodic details during narrative retelling, rate their memories as more significant and emotional (El Haj et al., 2016). Participants may thus provide more semantic, contextual, and metacognitive information to reflect how the particular event fits within their personal semantic landscape (Bluck & Habermas, 2000; Prebble et al., 2013; Addis & Tippett, 2008; Thomsen, 2009). Events that are rated more highly on these self-report measures may be linked to more life themes and knowledge structures that are then shared during the interview. It has been proposed that memories become more semanticized with more rehearsal (Skowronski & Walker, 2004). Consistent with prior work (Fuentes & Desrocher, 2013), external detail count was associated with self-reported rehearsal. Of note, proportion and density scores were unrelated to AM rehearsal. This suggests that proportion and density scores may be more robust to the effects of cuing accessible events versus events that were previously encoded but not previously retrieved (e.g. Svoboda & Levine, 2009).

Substantial debate remains regarding the role of self-reported memory vividness, and its relationship to performance-based AM. As an inherently phenomenological experience, researchers do not have direct access to the recollections of others. Participant reports of vividness have been related to medial temporal lobe activity and connectivity during recollection (Addis et al., 2004; Addis et al., 2011; Furman et al. 2012, Sheldon & Levine 2013; Thakral et al., 2020). However, here we found that vividness was not specifically related to internal count, showing a similar association with external count. In contrast, proportion and density scores showed no relationship with vividness. These observations, and those of others (Clark & Maguire, 2020; Setton et al., 2021), indicate that the AI does not capture many of the multi-faceted phenomenological features of autobiographical recollection.

Future work is needed to more deeply assess the reliability of the AI. Within the same testing session, reliability is high for events sampled from different points in time. Less is known about how test-retest reliability varies over more extended durations. To our knowledge, only a few studies have examined the persistence of AM recollection over time. One study with 12 middle aged adults found that repeated retrieval, more than the passage of time, increased the total of internal and external count (Nadel et al., 2007). The same group also reported that passage of one year led to increased detail recollection (Campbell et al., 2011). A more recent study in 16 younger adults found that internal count was nearly identical across two visits separated by 8 months, with only emotion/thought details receding over time (Barry et al., 2020). Replication in larger, more developmentally diverse samples is needed. Importantly, it is currently unknown

how AM recollection is affected by cueing, which may vary as a function of highly novel unrehearsed events versus more accessible and general events (for discussion, Armson et al., 2021). Additionally, it has been observed that internal details can be selectively augmented with a specificity induction (Madore et al., 2014). It is unknown if there is variable susceptibility to specificity induction across participants, and induction's relationship to stable trait estimation of AM.

The Autobiographical Interview is associated with age and gender

Consistent with the initial validation and many subsequent studies, we replicated established age effects in AM (e.g. Addis et al., 2008; Addis et al., 2010; Piolino et al., 2002; Piolino et al., 2006; St. Jacques & Levine, 2007): younger adults provided more internal details and older adults provided more external and semantic details. This typifies a more general trend in cognitive aging wherein fluid and cognitive control abilities decline while crystallized cognition is maintained or enhanced (Craik & Bialystok, 2006; Park et al., 2001; Park & Reuter-Lorenz, 2009; Spreng & Turner, 2019). However, non-mnemonic factors may also explain these age-related differences in performance on the AI. Differences in narrative style and discourse goals have been observed between younger and older adults (Adams et al., 1997; Carstensen et al., 1999; James et al., 1998). Similar patterns of age differences in reporting episodic and semantic details have been observed on a perceptual task, in the absence of memory demands (Gaesser et al., 2011). To begin to disentangle memory from narrative style, we previously introduced AI measures corrected for verbosity (Spreng et al., 2018). Here we find that the observed age differences in the AI extended to, and were even larger for, verbosity-corrected measures (proportion score, internal density, external density, and semantic density) indicating that AI age differences are not due to differences in number of words spoken. As we discuss below, this also suggests that proportion score and density measures may be more sensitive to age differences than detail count measures.

Previous findings of gender effects on the AI are mixed, with studies reporting gender differences (Fuentes and Desrocher, 2013), while other studies report no sex differences (Compère et al., 2016). Here we observed higher internal density scores for women, although men were more verbose overall when describing their memories. Previous research on gender differences in AM narrative length has been inconsistent (Grysman & Hudson, 2013). Gender differences have been reported (Niedźwieńska, 2003), although these differences may depend on the degree of retrieval support (Fuentes & Desrocher, 2013). Consistent with our findings, longer memory narratives have been reported in men (Grysman & Denney, 2017), including more off-topic statements (Baron & Bluck, 2009). However, adolescent and young adult women have also been observed to provide longer narratives (Bohn & Berntsen, 2008; Bohanek & Fivush, 2010; Fivush & Schwarzmueller, 1998; Grysman, 2017; Pillemer et al., 2003). In line with our findings of greater internal density in women, there is evidence that women elaborate further and provide more specific details during AM recollection (Bloise & Johnson, 2007; Grysman, 2017; Grysman et al., 2016; Grysman et al., 2020; Hayne & MacDonald, 2003; Seidlitz & Diener, 1998; Fivush et al., 2012).

Autobiographical Interview demonstrates good convergent validity

A primary interest of the present work was to evaluate the convergent validity of the AI by examining its relationship with other theoretically related cognitive constructs. Central to this validation, we predicted a positive association between internal AI metrics and episodic memory. Consistent with our predictions, we found that episodic memory was positively associated with all three internal measures across participants: internal count, proportion and internal density. This convergent validity provides strong support for the AI internal measures, as they track with laboratory-based episodic memory performance (even though internal count is not positively associated with proportion and internal density scores). Notably, internal density was robustly associated with the episodic index score in older adults, but not in the younger sample alone who tended to perform very well with less variability across subjects (although internal density was positively associated with the specific VPA recognition task in both samples). The association between episodic memory performance in the laboratory and AI scores was low (approximately .2). The low magnitude of this effect indicates that exploration of the inter-relationship between episodic and autobiographical memory requires samples larger than 100 to detect an association, and an order of magnitude higher to determine their dependencies in healthy adults. It is possible that episodic memory deficits may more tightly couple broader episodic and internal AM performance. We also found that episodic memory was negatively correlated with external density across the full sample and separately within younger and older adults. This finding supports the conceptualization of external density as a measure of non-episodic information produced in the absence of episodic richness.

Executive function is necessary for strategic memory retrieval (Conway & Pleydell-Pearce, 2000; Moscovitch, 1992), staying on task (Garon et al., 2008; Fisher & Kloos, 2016), and inhibiting irrelevant information (Ploner et al., 2001; Gazzaley et al., 2005). We predicted that executive function would be positively associated with internal measures, and negatively associated with external measures. We observed a small yet significant association between internal density and executive function, but no other significant associations emerged. This association was largely driven by reading span, which was positively correlated with both proportion and internal density measures. More so than associations with episodic memory, effect sizes relating the more distant constructs of executive function and AM were very small in our sample of healthy adults.

AM comprises both episodic and semantic features, and the AI aims to capture both with internal and external metrics. Thus, we assessed the relation between AI measures and semantic memory estimated by vocabulary size, a standard measure of crystalized cognition. Although we predicted that external AI measures would be positively associated with our semantic memory index, no associations were found. We contend that this is not a weakness of the AI, but rather a caution for how to interpret the external detail categories, which span a wide variety of representational knowledge and personal semantics, both of which are prominent during AM recollection (Renoult et al., 2012; Renoult et al., 2016). Novel scoring procedures for the AI have been introduced to better characterize semantic memory (e.g., Strikwerda-Brown et al., 2019; Renoult et al.,

2020), which may demonstrate clearer links between particular semantic sub-categories and the multi-faceted construct of semantic knowledge. Furthermore, the AI administration procedure requires participants to generate specific past events with a focus on recalling internal details. The production of non-specific event information is discouraged. In line with this idea, recent work has found that when participants are asked to describe opinions about everyday events, there is a rise in the number of external details and a reduction in the number of internal details, suggesting that AI details shift as a function of task demands (Strikwerda-Brown, Lévesque, Brambati & Sheldon, 2021). In the absence of any specific cues, individuals may default to schema-based interpretation and reconstruction of events: general impressions that are formed after events tend to be remembered first and serve as the basis for additional remembering (Bartlett, 1932). Better measurement of semantic AM may therefore require instructions with less of an episodic focus and/or more of an impressions-based focus (e.g., Madore et al., 2014).

The Remember/Know task (Tulving, 1972, 1985; Yonelinas et al.,1995) is a laboratory measure that aims to dissociate two types of recognition memory processes: recollection and familiarity. Recollection is thought to reflect more qualitative, episodic, high confidence remembering whereas familiarity is thought to reflect the subjective experience of 'knowing' in the absence of specific remembering (see Yonelinas, 2002 for review of dual-process models). In this assessment of convergent validity, the AI did not perform as predicted. No associations were observed for recollection, which were hypothesized to relate to internal measures. Contrary to a prior report (Piolino et al., 2006), familiarity was negatively associated with external measures. Participant insight into phenomenological aspects of recollection is a challenge to assess and may have been poorly characterized in this sample. Previous work has reported high rates of "Remember" responses for items that had not previously been seen (Odegard & Lampinen, 2004). Additionally, older adults could have answered inaccurately with "Remember" responses to compensate for memory deficits when memory-related age stereotypes were salient (Ryan & Campbell, 2021). Indeed, 29 older adults were excluded from analysis due to an insufficient number of "Know" responses. Additional work is needed to determine relations among self-reported recollection measures and the AI.

Autobiographical Interview is associated with subclinical depression, social cognition and personality

From our first aim assessing the reliability and validity of the AI, we determined that, overall, the AI performed well as a measure of AM. But AM is a multifaceted construct that permeates many aspects of daily life, affecting mood, decision-making, social interactions, and personality. Given the prevalence of AM capacities and previous studies suggesting the influence of these factors on AM, our second aim was to explore how the AI associated with these related, non-mnemonic constructs. In doing so, we drew on prior observations (often in smaller samples) to inform our predictions. Leveraging our large AI sample, we sought to replicate these findings and evaluate them as precedents in the literature in order to guide future work.

Over-general memory has been associated with clinical depression (Brittlebank et al., 1993; Hitchcock et al., 2014; Kuyken & Dalgleish, 1995; Liu et al., 2013; Williams & Broadbent, 1986; Williams & Scott, 1988; Williams et al., 2007; Wilson & Gregory, 2018). We assessed whether AI metrics would be sensitive to this effect in sub-clinical variance of depression symptomatology. In the younger adult sample, self-reported depression symptoms were negatively associated with proportion scores and positively associated with external measures of count and density. No associations were observed in older adults. These findings indicate that even minor depressive symptoms, below a diagnostic threshold, co-occur with less episodically rich AMs in young individuals. Notably, this may be driven more by the prevalence of external information.

The reconstructive nature of AM lends itself to a similar process by which to imagine and simulate future events (Schacter & Addis, 2007; Schacter, 2012). Indeed, a high correlation exists between internal details for past and future autobiographical events (Addis, Wong, & Schacter, 2008). This ability to envisage the future and mentally play out different scenarios, rooted in memory, has significant implications for decision-making (Boyer, 2008). For example, episodic simulation has been shown to reduce impulsive choices (Benoit et al., 2011; Peters & Büchel, 2010; Sasse et al., 2015). Such findings led to the prediction that individuals with richer episodic AM capacities are able to distance themselves from the temptation of rewards now by vividly visualizing and bringing themselves closer to their future selves (Gilbert & Wilson, 2007; Ersner-Hershfield et al., 2009; Bulley et al., 2016). Moreover, Lempert et al. (2020) reported a negative association between internal count and temporal discounting in 34 cognitively normal older adults, indicating that higher recollection of internal (specifically place, time, and perceptual) details was associated with more patient choices even into older age. In an attempt to replicate this effect in our well-powered sample, we found no associations between temporal discounting and AI metrics, including Lempert et al.'s perceptual/gist ratio. This observation warrants caution in linking internal AM to decision-making, consistent with prior reports of amnesic individuals performing within a normal range on a temporal discounting task (Kwan et al., 2015). That even semantic future thinking can minimize discounting in healthy controls (Palombo et al., 2016) further suggests that associations between future simulation and decision-making may not be specific to episodic processes.

Autobiographical recollection evokes similar patterns of brain activity as theory of mind reasoning (Buckner & Carroll, 2007). This original observation spurred much of the inquiry in understanding the relationship between AM and social cognition. For example, vivid recollection of helping another individual promotes the intention to engage in prosocial helping behavior (Gaesser & Schacter, 2014). We examined whether individual differences in empathy and theory of mind reasoning were associated with the AI metrics, predicting that more empathic individuals and those who are better at inferring emotional states would have more episodically rich memories. We observed a modest association between theory of mind, as assessed with the Reading the Mind in the Eyes task, and the AI proportion score. Scant associations were observed with trait empathy. Perhaps the induction of rich recollection or episodic simulation may be

necessary to elicit prosocial behavior (Gaesser & Schacter, 2014; Gaesser & Fowler, 2020), rather than trait level covariance. The modest number of associations observed reinforce the behavioral dissociation of social and mnemonic constructs.

Given that AM revolves around the self, we sought to examine whether self-reported personality traits were related to the AI metrics. Few associations were observed. Most notably, trait extraversion was positively associated with semantic density. More extraverted individuals are more likely to report using AM to fulfill social needs (Caci et al., 2019; Rasmussen & Berntsen, 2010; Webster, 1993), are more willing to share self-defining memories (McLean & Pasupathi, 2006), and prefer interpersonal reminiscence over other formats (Quackenbush & Barnett, 1995). Based on the social format of the AI, it is possible that more extraverted individuals share more contextual and framing details. Extraverted individuals also have greater rates of memory rehearsal, both generally and in conversation (Cappeliez & O'Rourke, 2002; Rasmussen & Berntsen, 2010; Rubin, Boals & Berntsen, 2008), suggesting that they may rehearse memories more frequently, contributing to higher rates of semantic information. Overall, correlation coefficients relating AI to the big five personality factors (DeYoung et al., 2007) were very modest. Sufficiently powered samples are needed to examine relationships between personality and AM, but are difficult to obtain given the current lengthy form of AI scoring. Advances in the scoring procedure that draw on machine learning (e.g., Peters et al., 2017) will be fruitful for this purpose.

Limited evidence for a two-factor structure in the Autobiographical Interview

In the preceding discussion, we assumed the validity of dissociating internal from external detail categories. The AI was designed to quantify and dissociate episodic from the non-episodic and semantic features of AM recall. An explicit test of the two-factor model with a CFA yielded poor model fit. Neither count nor density models outperformed a null model, and both showed large differences in residuals between the hypothesized and observed models. The CFA with density scores successfully dissociated internal and external components, but its structure was not robust. This suggests that there are nuanced relationships between sub-categories that the internal/external structure of the AI does not capture. Poor model performance may stem from the inclusion of semantic, repetition, and other details, which demonstrated particularly low regression coefficients in the CFA with density. This is supported by theory as semantic, repetition, and other detail categories are the only external sub-categories that do not describe a spatiotemporally specific event.

Based on the limitations of the internal/external model, we recommend that future work consider examining sub-categories in addition to overall internal and external variables when feasible. Future evaluation of alternative models to the two-factor AI, potentially separating semantic, repetition, and other categories from the external category, may be worthwhile. Exploratory data reduction of the detail category counts may reveal a latent structure that informs alternate scoring, but this would require further psychometric evaluation and replication. It is inconclusive whether the poor model fit informs ongoing debate regarding the dissociation or inter-relation of episodic and semantic memory

systems (e.g., Irish & Piguet, 2013; Renoult et al., 2019). However, our data strongly suggest that internal and external count scores should not be framed as distinct approximations of episodic and semantic memory given their interrelation. Instead, measures controlling for verbosity, such as density, may provide better estimates of dissociable AM components.

## Verbosity impacts the psychometric properties of the Autobiographical Interview

Across the psychometric assessment, the AI was broadly found to be a reliable and valid measure of AM. However, notable differences emerged as a function of scoring. Original count metrics for internal and external details were highly interrelated, with a shared dependency driven by the number of words spoken. These measures were also significantly linked to memory recency, as well as self-reported rehearsal.

Recently introduced density measures provide several strengths. Density measures significantly dissociate internal from external details, both in simple associations between AI metrics and the multivariate CFA. This distinction is essential for quantifying episodic AM. Density measures were similar across memory age, unrelated to memory rehearsal, and demonstrated larger effect sizes when comparing younger and older adults. In addition to being associated with episodic memory, internal density was also modestly correlated with executive function. External density also uniquely demonstrated negative associations with episodic memory, and positive associations with time integration (a measure of global contextual integration). Together, these observations suggest that density estimates may be an appropriate metric for individual differences in AM.

## Conclusions

Publication of the AI (Levine et al., 2002) was ground-breaking, offering a novel way for researchers to distill a naturalistic, qualitative measure of memory into discrete quantities of episodic and semantic recollection. Although nearly 20 years have passed before a comprehensive psychometric assessment, the AI is a highly reliable and valid measure of AM. By deriving and testing several AI metrics, here we find that verbosity-controlled variables may be better suited for detecting individual differences that reflect recollective processes beyond narrative style. Several open questions remain, such as the appropriate factor structure for the AI. Updates to scoring protocols that consider alternate models and expedite the scoring process will be instrumental for filling in these gaps to systematically assess AM in ever larger samples.

**Acknowledgements**
This project was supported by the Canadian Institute of Health Research and the Natural Sciences and Engineering Research Council of Canada. R.N.S. is a Research Scholar supported by the Fonds de la Recherche du Quebec - Santé.

**CRediT author statement**
**Amber W. Lockrow:** Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Review & Editing, Visualization; **Roni Setton:** Data Curation, Writing - Review & Editing; **Karen A.P. Spreng**: Resources, Writing - Review & Editing; **Signy Sheldon**: Writing - Review & Editing; **Gary R. Turner:** Writing - Review & Editing, Supervision, Project administration, Funding acquisition; **R. Nathan Spreng:** Conceptualization, Methodology, Writing - Original Draft, Review & Editing, Supervision, Project administration, Funding acquisition

The authors have no relevant financial or non-financial interests to disclose.

## References


Abellán-Martínez, M., Castellanos López, M. Á., Delgado-Losada, M. L., Yubero, R., Paúl, N., & Unturbe, F. M. (2019). Executive Control on Memory Test Performance across Life: Test of Memory Strategies. *The Spanish Journal of Psychology*, *22*, E50. https://doi.org/10.1017/sjp.2019.47

Adams, C., Smith, M. C., Nyquist, L., & Perlmutter, M. (1997). Adult Age-Group Differences in Recall for the Literal and Interpretive Meanings of Narrative Text. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *52B*(4), P187–P195. https://doi.org/10.1093/geronb/52B.4.P187

Addis, D. R., Moscovitch, M., Crawley, A. P., & McAndrews, M. P. (2004). Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. *Hippocampus*, *14*(6), 752-762.

Addis, D. R., Musicaro, R., Pan, L., & Schacter, D. L. (2010). Episodic simulation of past and future events in older adults: Evidence from an experimental recombination task. *Psychology and Aging*, *25*(2), 369–376. https://doi.org/10.1037/a0017280

Addis, D. R., Roberts, R. P., & Schacter, D. L. (2011). Age-related neural changes in autobiographical remembering and imagining. *Neuropsychologia*, *49*(13), 3656-3669.

Addis, D. R., Sacchetti, D. C., Ally, B. A., Budson, A. E., & Schacter, D. L. (2009). Episodic simulation of future events is impaired in mild Alzheimer's disease. *Neuropsychologia*, *47*(12), 2660–2671. https://doi.org/10.1016/j.neuropsychologia.2009.05.018

Addis, D.R., & Tippett, L. J. (2008). The contributions of autobiographical memory to the content and continuity of identity a social-cognitive neuroscience approach. *Self Continuity Individual and Collective Perspectives*.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-Related Changes in the Episodic Simulation of Future Events. *Psychological Science*, *19*(1), 33–41. https://doi.org/10.1111/j.1467-9280.2008.02043.x

Adler, J. M., Wagner, J. W., & McAdams, D. P. (2007). Personality and the coherence of psychotherapy narratives. *Journal of Research in Personality*, *41*(6), 1179–1198. https://doi.org/10.1016/j.jrp.2007.02.006

Althouse, A. D. (2016). Adjust for Multiple Comparisons? It's Not That Simple. *The Annals of Thoracic Surgery*, *101*(5), 1644–1645. https://doi.org/10.1016/j.athoracsur.2015.11.024

Armson, M. J., Diamond, N. B., Levesque, L., Ryan, J. D., & Levine, B. (2021). Vividness of recollection is supported by eye movements in individuals with high, but not low trait autobiographical memory. *Cognition*, *206*, 104487.

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. New York, NY, US: Cambridge University Press.

Bastin, C., Feyers, D., Jedidi, H., Bahri, M. A., Degueldre, C., Lemaire, C., Collette, F., & Salmon, E. (2013). Episodic autobiographical memory in amnestic mild cognitive impairment: What are the neural correlates? *Human Brain Mapping*, *34*(8), 1811–1825. https://doi.org/10.1002/hbm.22032

Barry, D. N., Clark, I. A., & Maguire, E. A. (2020). The relationship between hippocampal subfield volumes and autobiographical memory persistence. *Hippocampus*, *31*(4), 362-374.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Baron, J. M., & Bluck, S. (2009). Autobiographical memory sharing in everyday life: Characteristics of a good story. *International Journal of Behavioral Development*, *33*(2), 105–117. https://doi.org/10.1177/0165025408098039

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory (BDI-II)*, *10*. Pearson.

Benoit, R. G., Gilbert, S. J., & Burgess, P. W. (2011). A Neural Mechanism Mediating the Impact of Episodic Prospection on Farsighted Decisions. *Journal of Neuroscience*, *31*(18), 6771–6779. https://doi.org/10.1523/JNEUROSCI.6559-10.2011

Berna, F., Schönknecht, P., Seidl, U., Toro, P., & Schröder, J. (2012). Episodic autobiographical memory in normal aging and mild cognitive impairment: A population-based study. *Psychiatry Research*, *200*(2–3), 807–812. https://doi.org/10.1016/j.psychres.2012.03.022

Blagov, P. S., & Singer, J. A. (2004). Four dimensions of self-defining memories (specificity, meaning, content, and affect) and their relationships to self-restraint, distress, and repressive defensiveness. *Journal of personality*, *72*(3), 481-511.

Bluck, S., & Habermas, T. (2000). The life story schema. *Motivation and emotion*, *24*(2), 121-147.

Boettiger, C. A., Mitchell, J. M., Tavares, V. C., Robertson, M., Joslyn, G., D'Esposito, M., & Fields, H. L. (2007). Immediate Reward Bias in Humans: Fronto-Parietal Networks and a Role for the Catechol-O-Methyltransferase 158Val/Val Genotype. *Journal of Neuroscience*, *27*(52), 14383–14391. https://doi.org/10.1523/JNEUROSCI.2551-07.2007

Bohanek, J. G., & Fivush, R. (2010). Personal narratives, well-being, and gender in adolescence. *Cognitive Development*, *25*(4), 368-379.

Bohn, A., & Berntsen, D. (2008). Life story development in childhood: the development of life story abilities and the acquisition of cultural life scripts from late middle childhood to adolescence. *Developmental Psychology*, *44*(4), 1135.

Borrini, G., Dall'Ora, P., Sala, S. D., Marinelli, L., & Spinnler, H. (1989). Autobiographical memory. Sensitivity to age and education of a standardized enquiry. *Psychological Medicine*, *19*(1), 215–224. https://doi.org/10.1017/S0033291700011181

Boyer, P. (2008). Evolutionary economics of mental time travel? *Trends in Cognitive Sciences*, *12*(6), 219–224. https://doi.org/10.1016/j.tics.2008.03.003

Brainerd, C. J., & Pressley, M. (Eds.). (2013). *Basic processes in memory development: Progress in cognitive development research*. Springer Science & Business Media.

Brittlebank, A. D., Scott, J., Mark, J., Williams, G., & Ferrier, I. N. (1993). Autobiographical Memory in Depression: State or Trait Marker? *British Journal of Psychiatry*, *162*(1), 118–121. https://doi.org/10.1192/bjp.162.1.118

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, *21*(2), 230-258.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). *The Brain's Default Network*: *Anatomy, Function, and Relevance to Disease*. *Annals of the New York Academy of Sciences*, *1124*(1), 1–38. https://doi.org/10.1196/annals.1440.011

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Bulley, A., Henry, J., & Suddendorf, T. (2016). Prospection and the present moment: The role of episodic foresight in intertemporal choices between immediate and delayed rewards. *Review of General Psychology*, *20*(1), 29-47.

Caci, B., Cardaci, M., & Miceli, S. (2019). Autobiographical memory, personality, and Facebook mementos. *Europe's Journal of Psychology*, *15*(3), 614–636. https://doi.org/10.5964/ejop.v15i3.1713

Campbell, J., Nadel, L., Duke, D., & Ryan, L. (2011). Remembering all that and then some: Recollection of autobiographical memories after a 1-year delay. *Memory*, *19*(4), 406-415.

Cappeliez, P., & O'Rourke, N. (2002). Personality traits and existential concerns as predictors of the functions of reminiscence in older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(2), P116-P123.

Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). A Theory of Socioemotional Selectivity. *American Psychologist*, 17.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284.

Clark, I. A., & Maguire, E. A. (2020). Do questionnaires reflect their purported cognitive functions?. *Cognition*, *195*, 104114.

Coelho, S., Guerreiro, M., Chester, C., Silva, D., Maroco, J., Paglieri, F., & de Mendonça, A. (2019). Mental time travel in mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, *41*(8), 845–855. https://doi.org/10.1080/13803395.2019.1632269

Cohen, J. (1988). Statistical power analysis for the social sciences.

Cole, S. N., Gill, N. C. L., Conway, M. A., & Morrison, C. M. (2012). Rapid communication: Mental time travel: Effects of trial duration on episodic and semantic content. *Quarterly Journal of Experimental Psychology*, *65*(12), 2288–2296. https://doi.org/10.1080/17470218.2012.740053

Compère, L., Sperduti, M., Gallarda, T., Anssens, A., Lion, S., Delhommeau, M., Martinelli, P., Devauchelle, A.-D., Oppenheim, C., & Piolino, P. (2016). Sex Differences in the Neural Correlates of Specific and General Autobiographical Memory. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00285

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, *107*(2), 261.

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, *10*(3), 131–138. https://doi.org/10.1016/j.tics.2006.01.007

Crovitz, H. F., & Schiffman, H. (1974). Frequency of episodic memories as a function of their age. *Bulletin of the Psychonomic Society*, *4*(5), 517-518.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*(4), 450-466.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

De Brigard, F., Giovanello, K. S., Stewart, G. W., Lockrow, A. W., O'Brien, M. M., & Spreng, R. N. (2016). Characterizing the subjective experience of episodic past, future, and counterfactual thinking in healthy younger and older adults. *Quarterly Journal of Experimental Psychology*, *69*(12), 2358–2375. https://doi.org/10.1080/17470218.2015.1115529

De Brigard, F., Rodriguez, D. C., & Montañés, P. (2017). Exploring the experience of episodic past, future, and counterfactual thinking in younger and older adults: A study of a Colombian sample. *Consciousness and cognition*, *51*, 258-267.

Dede, A. J., 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*, *113*(47), 13480-13485.

Dede, A. J. O., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2016). Autobiographical memory, future imagining, and the medial temporal lobe. *Proceedings of the National Academy of Sciences*, *113*(47), 13474–13479. https://doi.org/10.1073/pnas.1615864113

Devitt, A. L., Addis, D. R., & Schacter, D. L. (2017). Episodic and semantic content of memory and imagination: A multilevel analysis. *Memory & Cognition*, *45*(7), 1078–1094. https://doi.org/10.3758/s13421-017-0716-1

Devitt, A. L., & Schacter, D. L. (2020). Looking on the bright side: Aging and the impact of emotional future simulation on subsequent memory. *The Journals of Gerontology: Series B*, *75*(9), 1831-1840.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

DuPre, E., Luh, W. M., & Spreng, R. N. (2016). Multi-echo fMRI replication sample of autobiographical memory, prospection and theory of mind reasoning tasks. *Scientific data*, *3*(1), 1-9.

El Haj, M., Kapogiannis, D., & Antoine, P. (2016). Phenomenological reliving and visual imagery during autobiographical recall in Alzheimer's disease. *Journal of Alzheimer's Disease*, *52*(2), 421-431.

Ersner-Hershfield, H., Garton, M. T., Ballard, K., Samanez-Larkin, G. R., & Knutson, B. (2009). Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. *Judgment and Decision making*, *4*(4), 280.

Fisher, A., & Kloos, H. (2016). Development of selective sustained attention: The role of executive functions.

Fivush, R., Bohanek, J. G., Zaman, W., & Grapin, S. (2012). Gender Differences in Adolescents' Autobiographical Narratives. *Journal of Cognition and Development*, *13*(3), 295–319. https://doi.org/10.1080/15248372.2011.590787

Fivush, R., & Schwarzmueller, A. (1998). Children remember childhood: Implications for childhood amnesia. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *12*(5), 455-473.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state." *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Ford, J. H., Rubin, D. C., & Giovanello, K. S. (2014). Effects of task instruction on autobiographical memory specificity in young and older adults. *Memory*, *22*(6), 722-736.

Fuentes, A., & Desrocher, M. (2013). The effects of gender on the retrieval of episodic and semantic components of autobiographical memory. *Memory*, *21*(6), 619–632. https://doi.org/10.1080/09658211.2012.744423

Furman, O., Mendelsohn, A., & Dudai, Y. (2012). The episodic engram transformed: Time reduces retrieval-related brain activity but correlates it with memory accuracy. *Learning & Memory*, *19*(12), 575-587.

Gaesser, B. (2013). Constructing memory, imagination, and empathy: a cognitive neuroscience perspective. *Frontiers in psychology*, *3*, 576.

Gaesser, B., & Fowler, Z. (2020). Episodic simulation of prosocial interaction: Investigating the roles of memory and imagination in facilitating a willingness to help others. *Psychology of Consciousness: Theory, Research, and Practice*, *7*(4), 376.

Gaesser, B., Sacchetti, D. C., Addis, D. R., & Schacter, D. L. (2011). Characterizing age-related changes in remembering the past and imagining the future. *Psychology and Aging*, *26*(1), 80–84. https://doi.org/10.1037/a0021054

Gaesser, B., & Schacter, D. L. (2014). Episodic simulation and episodic memory can increase intentions to help others. *Proceedings of the National Academy of Sciences*, *111*(12), 4415-4420.

Gamboz, N., De Vito, S., Brandimonte, M. A., Pappalardo, S., Galeone, F., Iavarone, A., & Sala, S. D. (2010). Episodic future thinking in amnesic mild cognitive impairment. *Neuropsychologia*, *48*(7), 2091–2097. https://doi.org/10.1016/j.neuropsychologia.2010.03.030

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological bulletin*, *134*(1), 31.

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of cognitive neuroscience*, *17*(3), 507-517.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science*, *317*(5843), 1351-1354.

Gilboa, A., Winocur, G., Rosenbaum, R. S., Poreh, A., Gao, F., Black, S. E., Westmacott, R., & Moscovitch, M. (2006). Hippocampal contributions to recollection in retrograde and anterograde amnesia. *Hippocampus*, *16*(11), 966–980. https://doi.org/10.1002/hipo.20226

Grysman, A. (2017). Gender differences in episodic encoding of autobiographical memory. *Journal of Applied Research in Memory and Cognition*, *6*(1), 51-59.

Grysman, A., & Denney, A. (2017). Gender, experimenter gender and medium of report influence the content of autobiographical memory report. *Memory*, *25*(1), 132–145. https://doi.org/10.1080/09658211.2015.1133829

Grysman, A., Fivush, R., Merrill, N. A., & Graci, M. (2016). The influence of gender and gender typicality on autobiographical memory across event types and age groups. *Memory & Cognition*, *44*(6), 856–868. https://doi.org/10.3758/s13421-016-0610-2

Grysman, A., Harris, C. B., Barnier, A. J., & Savage, G. (2020). Long-married couples recall their wedding day: the influence of collaboration and gender on autobiographical memory recall. *Memory*, *28*(1), 18-33.

Grysman, A., & Hudson, J. A. (2013). Gender differences in autobiographical memory: Developmental and methodological considerations. *Developmental Review*, *33*(3), 239-272.

Hayne, H., & MacDonald, S. (2003). The socialization of autobiographical memory in children and adults: The roles of culture and gender. *Autobiographical memory and the construction of a narrative self: Developmental and cultural perspectives*, 99-120.

Hilverman, C., Cook, S. W., & Duff, M. C. (2016). Hippocampal declarative memory supports gesture production: Evidence from amnesia. *Cortex*, *85*, 25–36. https://doi.org/10.1016/j.cortex.2016.09.015

Hitchcock, C., Nixon, R. D., & Weber, N. (2014). A review of overgeneral memory in child psychopathology. *British Journal of Clinical Psychology*, *53*(2), 170-193.

Kirwan, C. B., Bayley, P. J., Galvan, V. V., & Squire, L. R. (2008). Detailed recollection of remote autobiographical memory after damage to the medial temporal lobe. *Proceedings of the National Academy of Sciences*, *105*(7), 2676–2680. https://doi.org/10.1073/pnas.0712155105

Kuyken, W., & Dalgleish, T. (1995). Autobiographical memory and depression. *British journal of clinical psychology*, *34*(1), 89-92.

Hooper, D., Coughlan, J. and Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods, 6*(1), 53 – 60.

Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: Evidence from semantic dementia. *Brain*, *135*(7), 2178–2191. https://doi.org/10.1093/brain/aws119

Irish, M., & Piguet, O. (2013). The Pivotal Role of Semantic Memory in Remembering the Past and Imagining the Future. *Frontiers in Behavioral Neuroscience*, *7*. https://doi.org/10.3389/fnbeh.2013.00027

James, L. E., Burke, D. M., Austin, A., & Hulme, E. (1998). Production and perception of" verbosity" in younger and older adults. *Psychology and aging*, *13*(3), 355.

Judd, C., Westfall, J., & Kenny, D. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. Journal of Personality and Social Psychology, 103(1), 54-69. https://doi.org/10.1037/a0028347

Kopelman, M. D., Baddeley, A. D., & Wilson, B. A. (1990). *AMI: The Autobiographical Memory Interview: Manual*. Harcourt Assessment.

Kopelman, M. D., Wilson, B. A., & Baddeley, A. D. (1989). The autobiographical memory interview: A new assessment of autobiographical and personal semantic memory in amnesic patients. *Journal of Clinical and Experimental Neuropsychology*, *11*(5), 724–744. https://doi.org/10.1080/01688638908400928

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*(1), 1-26.

Kwan, D., Carson, N., Addis, D. R., & Rosenbaum, R. S. (2010). Deficits in past remembering extend to future imagining in a case of developmental amnesia. *Neuropsychologia*, *48*(11), 3179–3186. https://doi.org/10.1016/j.neuropsychologia.2010.06.011

Kwan, D., Craver, C. F., Green, L., Myerson, J., Gao, F., Black, S. E., & Rosenbaum, R. S. (2015). Cueing the personal future to reduce discounting in intertemporal choice: Is episodic prospection necessary?. *Hippocampus*, *25*(4), 432-443.

Kwan, D., Kurczek, J., & Rosenbaum, R. S. (2016). Specific, personally meaningful cues can benefit episodic prospection in medial temporal lobe amnesia. *British Journal of Clinical Psychology*, *55*(2), 137–153. https://doi.org/10.1111/bjc.12095

Lempert, K. M., MacNear, K. A., Wolk, D. A., & Kable, J. W. (2020). Links between autobiographical memory richness and temporal discounting in older adults. *Scientific Reports*, *10*(1), 6431. https://doi.org/10.1038/s41598-020-63373-1

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

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Linton, M. (1975). Memory for real-world events. *Explorations in cognition*, 376-404.

Liu, X., Li, L., Xiao, J., Yang, J., & Jiang, X. (2013). Abnormalities of autobiographical memory of patients with depressive disorders: A meta-analysis. *Psychology and Psychotherapy: Theory, Research and Practice*, *86*(4), 353–373. https://doi.org/10.1111/j.2044-8341.2012.02077.x

Löckenhoff, C. E., O'Donoghue, T., & Dunning, D. (2011). Age differences in temporal discounting: the role of dispositional affect and anticipated emotions. *Psychology and aging*, *26*(2), 274.

Madore, K. P., Gaesser, B., & Schacter, D. L. (2014). Constructive episodic simulation: dissociable effects of a specificity induction on remembering, imagining, and describing in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 609.

McAdams, D. P., Anyidoho, N. A., Brown, C., Huang, Y. T., Kaplan, B., & Machado, M. A. (2004). Traits and Stories: Links Between Dispositional and Narrative Features of Personality. *Journal of Personality*, *72*(4), 761–784. https://doi.org/10.1111/j.0022-3506.2004.00279.x

McLean, K. C., & Pasupathi, M. (2006). Collaborative narration of the past and extraversion. *Journal of Research in Personality*, *40*(6), 1219–1231. https://doi.org/10.1016/j.jrp.2005.11.006

Meulenbroek, O., Rijpkema, M., Kessels, R. P. C., Rikkert, M. G. M. O., & Fernández, G. (2010). Autobiographical memory retrieval in patients with Alzheimer's disease. *NeuroImage*, *53*(1), 331–340. https://doi.org/10.1016/j.neuroimage.2010.05.082

Miller, T. D., Chong, T. T.-J., Aimola Davies, A. M., Johnson, M. R., Irani, S. R., Husain, M., Ng, T. W., Jacob, S., Maddison, P., Kennard, C., Gowland, P. A., & Rosenthal, C. R. (2020). Human hippocampal CA3 damage disrupts both recent and remote episodic memories. *ELife*, *9*, e41836. https://doi.org/10.7554/eLife.41836

Miloyan, B., McFarlane, K., & Vásquez-Echeverría, A. (2019). The adapted Autobiographical interview: A systematic review and proposal for conduct and reporting. *Behavioural brain research*, *370*, 111881.

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of cognitive neuroscience*, *4*(3), 257-267.

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur, G., & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: A unified account based on multiple trace theory: A unified account based on multiple trace theory. *Journal of Anatomy*, *207*(1), 35–66. https://doi.org/10.1111/j.1469-7580.2005.00421.x

Murphy, K. J., Troyer, A. K., Levine, B., & Moscovitch, M. (2008). Episodic, but not semantic, autobiographical memory is reduced in amnestic mild cognitive impairment. *Neuropsychologia*, *46*(13), 3116–3123. https://doi.org/10.1016/j.neuropsychologia.2008.07.004

Murre, J. M. J., Kristo, G., & Janssen, S. M. J. (2014). The effect of self-reported habitual sleep quality and sleep length on autobiographical memory. *Memory*, *22*(6), 633–645. https://doi.org/10.1080/09658211.2013.811253

Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the experimental analysis of behavior*, *76*(2), 235-243.

Nadel, L., Campbell, J., & Ryan, L. (2007). Autobiographical memory retrieval and hippocampal activation as a function of repetition and the passage of time. *Neural plasticity*, *2007*.

Niedźwieńska, A. (2003). Gender differences in vivid memories. *Sex Roles*, *49*(7), 321-331.

Odegard, T., & Lampinen, J. (2004). Memory conjunction errors for autobiographical events: More than just familiarity. *Memory*, *12*(3), 288–300. https://doi.org/10.1080/09658210244000621

Orwin, R. G. (1994). Evaluation coding decisions. In H. H. Cooper, L.V. (Ed.), *The handbook of research synthesis* . New York: Russel Sage Foundation.

Palombo, D. J., Keane, M. M., & Verfaellie, M. (2016). Using future thinking to reduce temporal discounting: Under what circumstances are the medial temporal lobes critical?. *Neuropsychologia*, *89*, 437-444.

Palombo, D. J., Sheldon, S., & Levine, B. (2018). Individual Differences in Autobiographical Memory. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2018.04.007

Park, D. C., Polk, T. A., Mikels, J. A., Taylor, S. F., & Marshuetz, C. (2001). Cerebral aging: Integration of brain and behavioral models of cognitive function. *Dialogues in Clinical Neuroscience*, *3*(3), 15.

Park, D. C., & Reuter-Lorenz, P. (2009). The Adaptive Brain: Aging and Neurocognitive Scaffolding. *Annual Review of Psychology*, *60*(1), 173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656

Peters, J., & Büchel, C. (2010). Episodic Future Thinking Reduces Reward Delay Discounting through an Enhancement of Prefrontal-Mediotemporal Interactions. *Neuron*, *66*(1), 138–148. https://doi.org/10.1016/j.neuron.2010.03.026

Peters, S. L., Fan, C. L., & Sheldon, S. (2019). Episodic memory contributions to autobiographical memory and open-ended problem-solving specificity in younger and older adults. *Memory & cognition*, *47*(8), 1592-1605.

Peters, J., Wiehler, A., & Bromberg, U. (2017). Quantitative text feature analysis of autobiographical interview data: prediction of episodic details, semantic details and temporal discounting. *Scientific reports*, *7*(1), 1-13.

Pillemer, D., Wink, P., DiDonato, T., & Sanborn, R. (2003). Gender differences in autobiographical memory styles of older adults. *Memory*, *11*(6), 525–532. https://doi.org/10.1080/09658210244000117

Piolino, P., Desgranges, B., Benali, K., & Eustache, F. (2002). Episodic and semantic remote autobiographical memory in ageing. *Memory*, *10*(4), 239–257. https://doi.org/10.1080/09658210143000353

Piolino, P., Desgranges, B., Clarys, D., Guillery-Girard, B., Taconnat, L., Isingrini, M., & Eustache, F. (2006). Autobiographical memory, autonoetic consciousness, and self-perspective in aging. *Psychology and Aging*, *21*(3), 510–525. https://doi.org/10.1037/0882-7974.21.3.510

Ploner, C.J., Ostendorf, F., Brandt, S.A., Gaymard, B.M., Rivaud-Péchoux S., & Ploner M. (2001). Behavioural relevance modulates access to spatial working memory in humans. *European Journal of Neuroscience*, *13*(2), 357-363.

Prebble, S. C., Addis, D. R., & Tippett, L. J. (2013). Autobiographical memory and sense of self. *Psychological Bulletin*, *139*(4), 815–840. https://doi.org/10.1037/a0030146

Quackenbush, S. W., & Barnett, M. A. (1995). Correlates of Reminiscence Activity among Elderly Individuals. *The International Journal of Aging and Human Development*, *41*(2), 169–181. https://doi.org/10.2190/EFW2-BEK9-0G1Y-NP2R

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rabin, J. S., Gilboa, A., Stuss, D. T., Mar, R. A., & Rosenbaum, R. S. (2010). Common and Unique Neural Correlates of Autobiographical Memory and Theory of Mind. *Journal of Cognitive Neuroscience*, *22*(6), 1095–1111. https://doi.org/10.1162/jocn.2009.21344

Race, E., Keane, M. M., & Verfaellie, M. (2011). Medial Temporal Lobe Damage Causes Deficits in Episodic Memory and Episodic Future Thinking Not Attributable to Deficits in Narrative Construction. *Journal of Neuroscience*, *31*(28), 10262–10269. https://doi.org/10.1523/JNEUROSCI.1145-11.2011

Rasmussen, A. S., & Berntsen, D. (2010). Personality traits and autobiographical memory: Openness is positively related to the experience and usage of recollections. *Memory*, *18*(7), 774–786. https://doi.org/10.1080/09658211.2010.514270

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and motor skills*, *8*(3), 271-276.

Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: At the crossroads of semantic and episodic memory. *Trends in Cognitive Sciences*, *16*(11), 550–558. https://doi.org/10.1016/j.tics.2012.09.003

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From Knowing to Remembering: The Semantic–Episodic Distinction. *Trends in Cognitive Sciences*, *23*(12), 1041–1057. https://doi.org/10.1016/j.tics.2019.09.008

Renoult, L., Tanguay, A., Beaudry, M., Tavakoli, P., Rabipour, S., Campbell, K., Moscovitch, M., Levine, B., & Davidson, P. S. R. (2016). Personal semantics: Is it distinct from episodic and semantic memory? An electrophysiological study of memory for autobiographical facts and repeated events in honor of Shlomo Bentin. *Neuropsychologia*, *83*, 242–256. https://doi.org/10.1016/j.neuropsychologia.2015.08.013

Renoult, L., Armson, M. J., Diamond, N. B., Fan, C. L., Jeyakumar, N., Levesque, L., Oliva, L., McKinnon, M., Papadopoulos, A., Selarka, D., St Jacques, P. L., & Levine, B. (2020). Classification of general and personal semantic details in the Autobiographical Interview. *Neuropsychologia*, *144*, 107501. https://doi.org/10.1016/j.neuropsychologia.2020.107501

Revelle, W. (2012). Procedures for psychological, psychometric, and personality research. *Acesso em*, *9*.

Robin, J., & Moscovitch, M. (2017). Familiar real-world spatial cues provide memory benefits in older and younger adults. *Psychology and Aging*, *32*(3), 210.

Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., & Moscovitch, M. (2009). Amnesia as an impairment of detail generation and binding: Evidence from personal, fictional, and semantic narratives in K.C. *Neuropsychologia*, *47*(11), 2181–2187. https://doi.org/10.1016/j.neuropsychologia.2008.11.028

Rosenbaum, R. S., McKinnon, M. C., Levine, B., & Moscovitch, M. (2004). Visual imagery deficits, impaired strategic retrieval, or memory loss: Disentangling the nature of an amnesic person's autobiographical memory deficit. *Neuropsychologia*, *42*(12), 1619–1635. https://doi.org/10.1016/j.neuropsychologia.2004.04.010

Rosenbaum, R. S., Moscovitch, M., Foster, J. K., Schnyer, D. M., Gao, F., Kovacevic, N., Verfaellie, M., Black, S. E., & Levine, B. (2008). Patterns of Autobiographical Memory Loss in Medial-Temporal Lobe Amnesic Patients. *Journal of Cognitive Neuroscience*, *20*(8), 1490–1506. https://doi.org/10.1162/jocn.2008.20105

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, *48*(2), 1-36.

Rothman, K. J. (1990). No Adjustments Are Needed for Multiple Comparisons: *Epidemiology*, *1*(1), 43–46. https://doi.org/10.1097/00001648-199001000-00010

Rubin, D. C. (Ed.). (1988). *Autobiographical memory*. New York: Cambridge University Press.

Rubin, D. C., Boals, A., & Berntsen, D. (2008). Memory in posttraumatic stress disorder: properties of voluntary and involuntary, traumatic and nontraumatic autobiographical memories in people with and without posttraumatic stress disorder symptoms. *Journal of Experimental Psychology: General*, *137*(4), 591.

Rubin, D. C., & Siegler, I. C. (2004). Facets of personality and the phenomenology of autobiographical memory. *Applied Cognitive Psychology*, *18*(7), 913–930. https://doi.org/10.1002/acp.1038

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological review*, *103*(4), 734.

Ryan, A. D., & Campbell, K. L. (2021). The ironic effect of older adults' increased task motivation: Implications for neurocognitive aging. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01963-4

Sasse, L. K., Peters, J., Büchel, C., & Brassen, S. (2015). Effects of prospective thinking on intertemporal choice: The role of familiarity: Future Event Construction and Delay Discounting. *Human Brain Mapping*, *36*(10), 4210–4221. https://doi.org/10.1002/hbm.22912

Saville, D. J. (1990). Multiple Comparison Procedures: The Practical Solution. *The American Statistician*, *44*(2), 174–180. https://doi.org/10.1080/00031305.1990.10475712

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, *67*(8), 603.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 773–786. https://doi.org/10.1098/rstb.2007.2087

Schroots, J. J., Van Dijkum, C., & Assink, M. H. (2004). Autobiographical memory from a life span perspective. *The International Journal of Aging and Human Development*, *58*(1), 69-85.

Seidlitz, L., & Diener, E. (1998). Sex differences in the recall of affective experiences. *Journal of personality and social psychology*, *74*(1), 262.

Setton, R., Lockrow, A. W., Turner, G. R., & Spreng, R. N. (2021). Troubled past: A critical psychometric assessment of the self-report Survey of Autobiographical Memory (SAM). *Behavior Research Methods*, 1-26.

Sheldon, S., Vandermorris, S., Al-Haj, M., Cohen, S., Winocur, G., & Moscovitch, M. (2015). Ill-defined problem solving in amnestic mild cognitive impairment: Linking episodic memory to effective solution generation. *Neuropsychologia*, *68*, 168-175.

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*.

Sheldon, S., & Levine, B. (2013). Same as it ever was: Vividness modulates the similarities and differences between the neural networks that support retrieving remote and recent autobiographical memories. *NeuroImage*, *83*, 880-891.

Shipley, W. C., Gruber, C. P, Martin, T. A., & Klein, A. M. (2009). Shipley-2 manual. Los Angeles, CA: Western Psychological Services.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.

Skowronski, J. J., & Walker, W. R. (2004). How describing autobiographical events can affect autobiographical memories. *Social Cognition*, *22*(5: Special issue), 555-590.

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

Spreng, R. N., & Grady, C. L. (2010). Patterns of Brain Activity Supporting Autobiographical Memory, Prospection, and Theory of Mind, and Their Relationship to the Default Mode Network. *Journal of Cognitive Neuroscience*, *22*(6), 1112–1123. https://doi.org/10.1162/jocn.2009.21282

Spreng, R. N., Lockrow, A. W., DuPre, E., Setton, R., Spreng, K. A. P., & Turner, G. R. (2018). Semanticized autobiographical memory and the default – executive coupling hypothesis of aging. *Neuropsychologia*, *110*, 37–43. https://doi.org/10.1016/j.neuropsychologia.2017.06.009

Spreng, R. N., & Mar, R. A. (2012). I remember you: A role for memory in social cognition and the functional neuroanatomy of their interaction. *Brain Research*, *1428*, 43–50. https://doi.org/10.1016/j.brainres.2010.12.024

Spreng*, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale Development and Initial Validation of a Factor-Analytic Solution to Multiple Empathy Measures. *Journal of Personality Assessment*, *91*(1), 62–71. https://doi.org/10.1080/00223890802484381

Spreng, R. N., & Turner, G. R. (2019). The shifting architecture of cognition and brain function in older adulthood. *Perspectives on Psychological Science*, *14*(4), 523-542.

Squire, L. R., van der Horst, A. S., McDuff, S. G. R., Frascino, J. C., Hopkins, R. O., & Mauldin, K. N. (2010). Role of the hippocampus in remembering the past and imagining the future. *Proceedings of the National Academy of Sciences*, *107*(44), 19044–19048. https://doi.org/10.1073/pnas.1014391107

St. Jacques, P. L., & Levine, B. (2007). Ageing and autobiographical memory for emotional and neutral events. *Memory*, *15*(2), 129–144. https://doi.org/10.1080/09658210601119762

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

Steinvorth, S., Levine, B., & Corkin, S. (2005). Medial temporal lobe structures are needed to re-experience remote autobiographical memories: Evidence from H.M. and W.R. *Neuropsychologia*, *43*(4), 479–496. https://doi.org/10.1016/j.neuropsychologia.2005.01.001

Strikwerda-Brown, C., Williams, K., Lévesque, M., Brambati, S., & Sheldon, S. (2021). What are your thoughts? Exploring age-related changes in episodic and semantic autobiographical content on an open-ended retrieval task. *Memory*, 1-9.

Strikwerda-Brown, C., Mothakunnel, A., Hodges, J. R., Piguet, O., & Irish, M. (2019). External details revisited – A new taxonomy for coding 'non-episodic' content during autobiographical memory retrieval. *Journal of Neuropsychology*, *13*(3), 371–397. https://doi.org/10.1111/jnp.12160

Svoboda, E., & Levine, B. (2009). The effects of rehearsal on the functional neuroanatomy of episodic autobiographical and semantic remembering: a functional magnetic resonance imaging study. *Journal of neuroscience*, *29*(10), 3073-3082.

Terrett, G., Rose, N. S., Henry, J. D., Bailey, P. E., Altgassen, M., Phillips, L. H., ... & Rendell, P. G. (2016). The relationship between prospective memory and episodic future thinking in younger and older adulthood. *Quarterly Journal of Experimental Psychology*, *69*(2), 310-323.

Thakral, P. P., Madore, K. P., Addis, D. R., & Schacter, D. L. (2020). Reinstatement of event details during episodic simulation in the hippocampus. *Cerebral Cortex*, *30*(4), 2321-2337.

Thomsen, D. K. (2009). There is more to life stories than memories. *Memory*, *17*(4), 445-457.

Tulving, E. (1972). Organization of memory. *Episodic and semantic memory*.

Tulving, E. (1985). How Many Memory Systems Are There? *American Psychologist*, 14.

Vandermorris, S., Sheldon, S., Winocur, G., & Moscovitch, M. (2013). Differential Contributions of Executive and Episodic Memory Functions to Problem Solving in Younger and Older Adults. *Journal of the International Neuropsychological Society*, *19*(10), 1087–1096. https://doi.org/10.1017/S1355617713000982

Villanueva, R. A. M., & Chen, Z. J. (2019). ggplot2: elegant graphics for data analysis.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive psychology*, *18*(2), 225-252.

Webster, J. D. (1993). Construction and Validation of the Reminiscence Functions Scale. *Journal of Gerontology*, *48*(5), P256–P262. https://doi.org/10.1093/geronj/48.5.P256

Wechsler, D. (2008). WAIS-IV Administration and Scoring Manual. San Antonio, TX: The Psychological Corporation.

Wei, T., & Simko, V. (2017). R package "corrplot": Visualization of a Correlation Matrix.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Slotkin, J., Carlozzi, N. E., Bauer, P. J., Wallner-Allen, K., Fox, N., Havlik, R., Beaumont, J. L., Mungas, D., Manly, J. J., Moy, C., Conway, K., Edwards, E., Nowinski, C. J., & Gershon, R. (2014). The Cognition Battery of the NIH Toolbox for Assessment of Neurological and Behavioral Function: Validation in an Adult Sample.

*Journal of the International Neuropsychological Society*, *20*(6), 567–578. https://doi.org/10.1017/S1355617714000320

Wessel, I., Postma, I. R., Huntjens, R. J. C., Crane, C., Smets, J., Zeeman, G. G., & Barnhofer, T. (2014). Differential correlates of autobiographical memory specificity to affective and self-discrepant cues. *Memory*, *22*(6), 655–668. https://doi.org/10.1080/09658211.2013.811255

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin, 133*(1), 122-148. doi:10.1037/0033-2909.133.1.122

Williams, J. M., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of abnormal psychology*, *95*(2), 144.

Williams, J. M. G., & Scott, J. (1988). Autobiographical memory in depression. *Psychological Medicine*, *18*(3), 689–695. https://doi.org/10.1017/S0033291700008370

Wilson, F. C. L., & Gregory, J. D. (2018). Overgeneral autobiographical memory and depression in older adults: a systematic review. *Aging & mental health*, *22*(5), 575-586.

Wood, W. J., & Conway, M. (2006). Subjective impact, meaning making, and current and recalled emotions for self-defining memories. *Journal of personality*, *74*(3), 811-846.

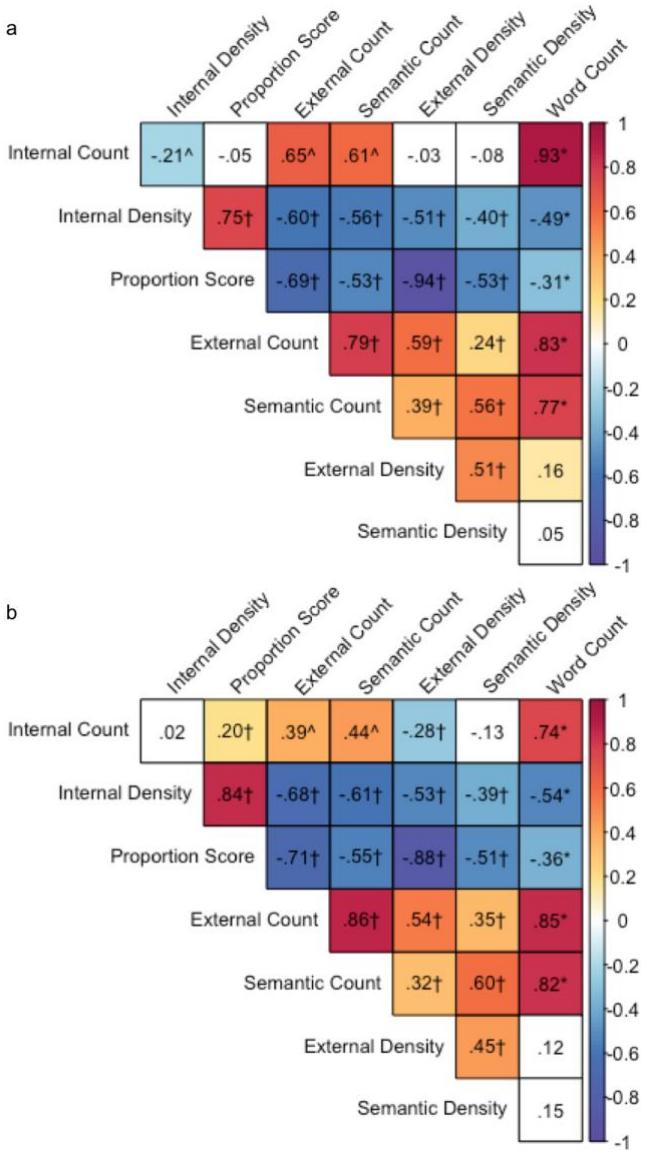
Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, *46*(3), 441-517.

Yonelinas, A. P., & Jacoby, L. L. (1995). The Relation between Remembering and Knowing as Bases for Recognition: Effects of Size Congruency. *Journal of Memory and Language*, *34*(5), 622–643. https://doi.org/10.1006/jmla.1995.1028

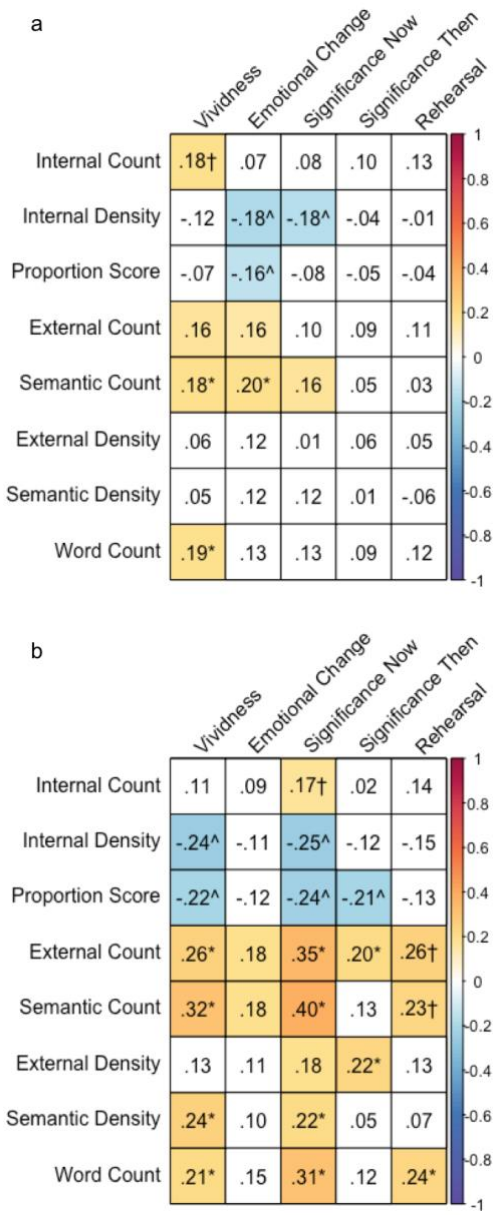
Yubero, R., Gil, P., Paul, N., & Maestú, F. (2011). Influence of memory strategies on memory test performance: A study in healthy and pathological aging. *Aging, Neuropsychology, and Cognition*, *18*(5), 497–515. https://doi.org/10.1080/13825585.2011.597840

Zavagnin, M., De Beni, R., Borella, E., & Carretti, B. (2016). Episodic future thinking: The role of working memory and inhibition on age-related differences. *Aging Clinical and Experimental Research*, *28*(1), 109–119. https://doi.org/10.1007/s40520-015-0368-6

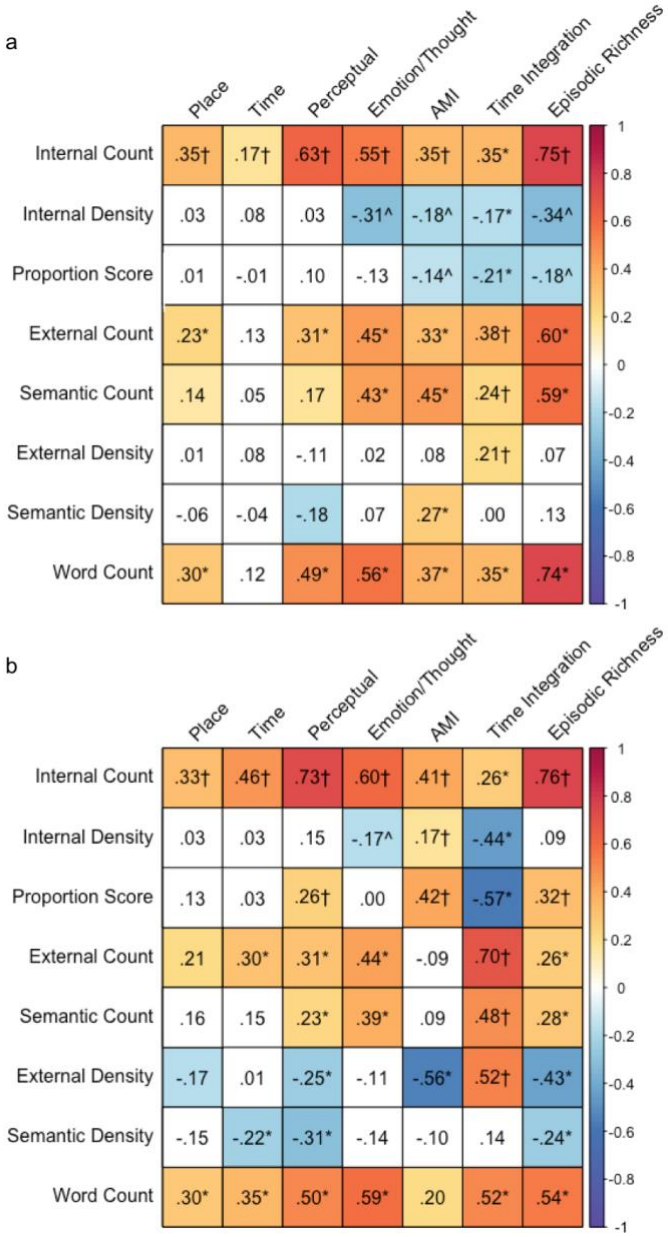
## Supplemental Materials



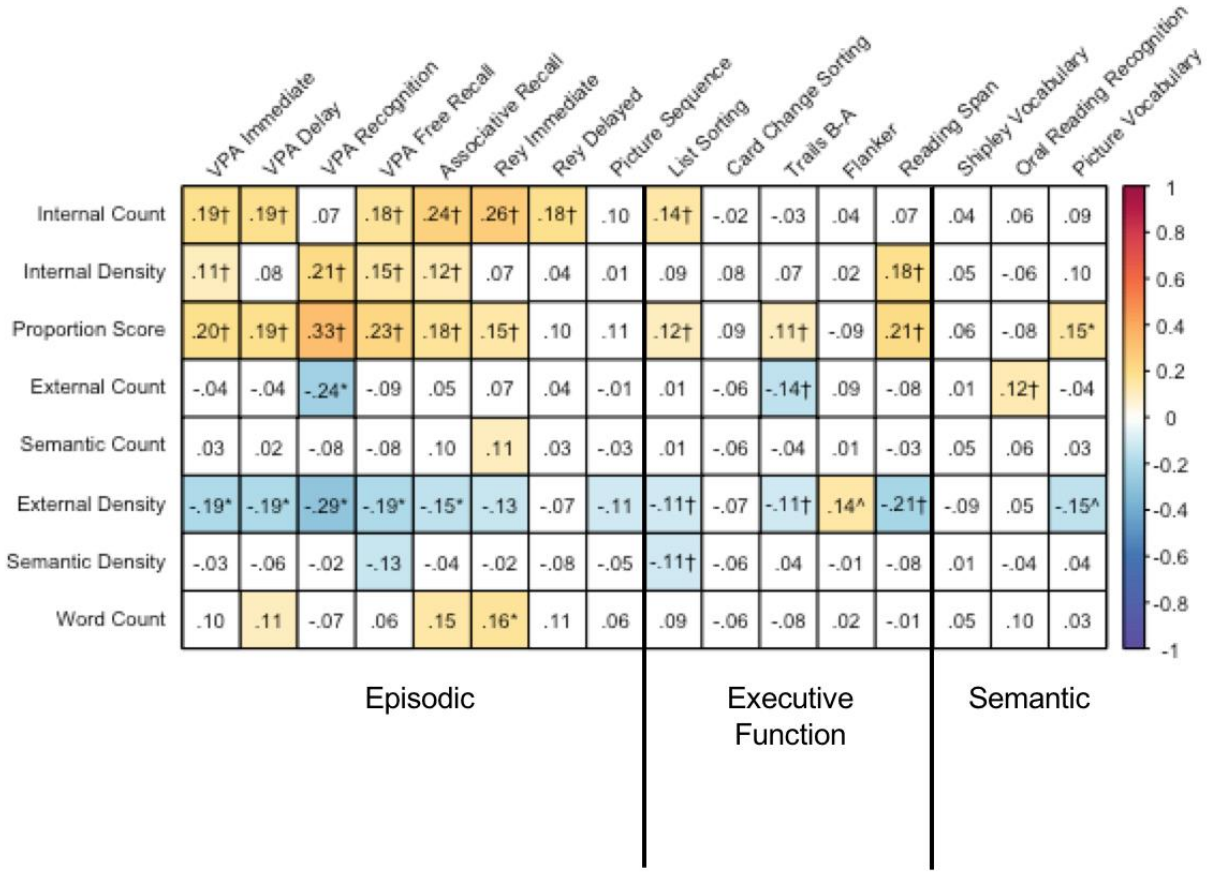

*Supplemental Figure 1.* Internal Consistency of the Autobiographical Interview Measures of Interest in Younger and Older Adults. Partial product-moment correlations controlling for gender between the 8 AI measures of interest in a) younger adults only b) older adults only. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction, ^ associations opposite to prediction ($p < .05$, uncorrected).



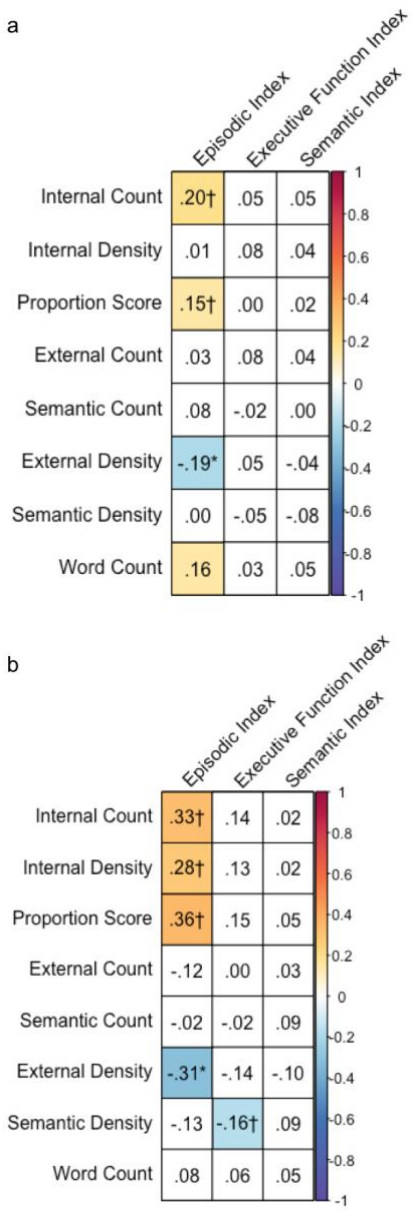

*Supplemental Figure 2.* Internal Consistency of the Autobiographical Interview Participant Self-Report Ratings in Younger and Older Adults. Partial Spearman correlations controlling for gender between the 8 AI measures of interest and self-report ratings controlling for gender in a) younger adults and b) older adults. The matrix contains partial ρ values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction, ^ associations opposite to prediction ($p < .05$, uncorrected).




*Supplemental Figure 3.* Internal Consistency of the Autobiographical Interview Scorer Ratings in Younger and Older Adults. AMI=Autobiographical Memory Interview. Partial Spearman Correlations between the 8 AI measures of interest and scorer overall ratings controlling for gender in a) younger adults only b) older adults only. The matrix contains partial ρ values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction, ^ associations opposite to prediction ($p < .05$, uncorrected).

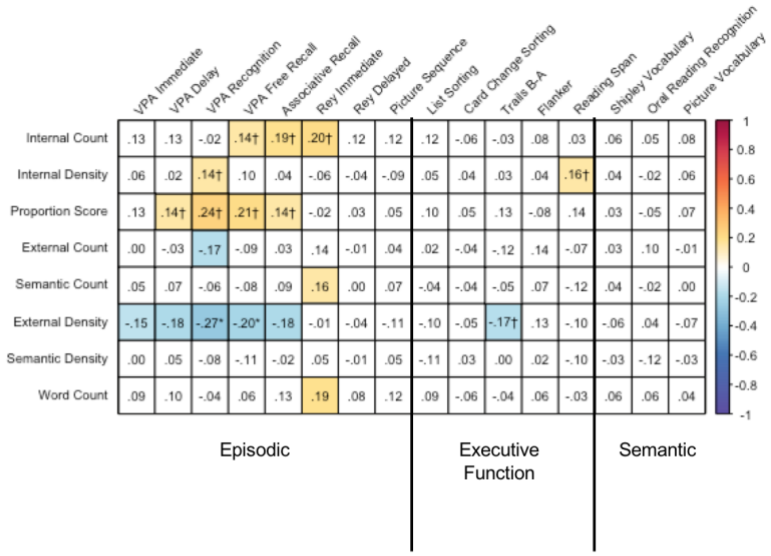



*Supplemental Figure 4.* Convergent Validity of the Autobiographical Interview with Episodic, Fluid, and Semantic Measures. Partial product-moment correlations controlling for age and gender across all participants between the 8 AI measures of interest and episodic, fluid, and semantic measures in all participants. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction.

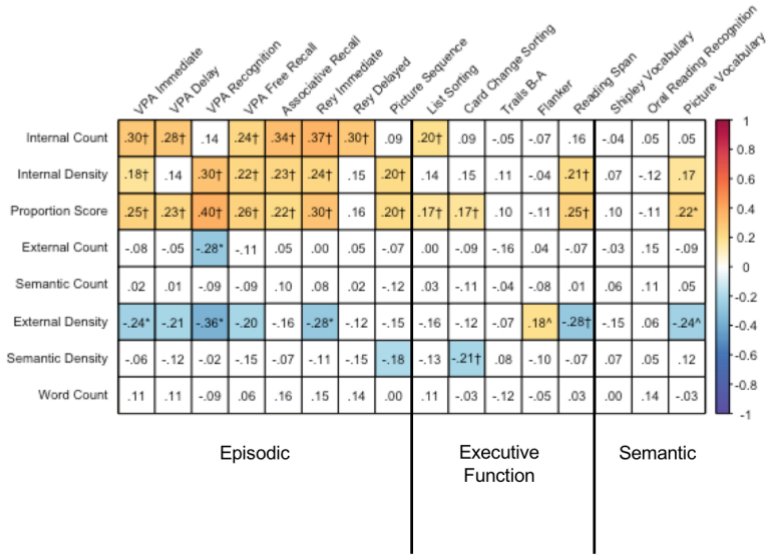



*Supplemental Figure 5.* Convergent Validity of the Autobiographical Interview with Episodic, Fluid, and Semantic Index Scores in Younger and Older Adults. Partial product-moment correlations controlling for gender between index scores and the 8 AI measures of interest in a) younger adults and b) older adults. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction.

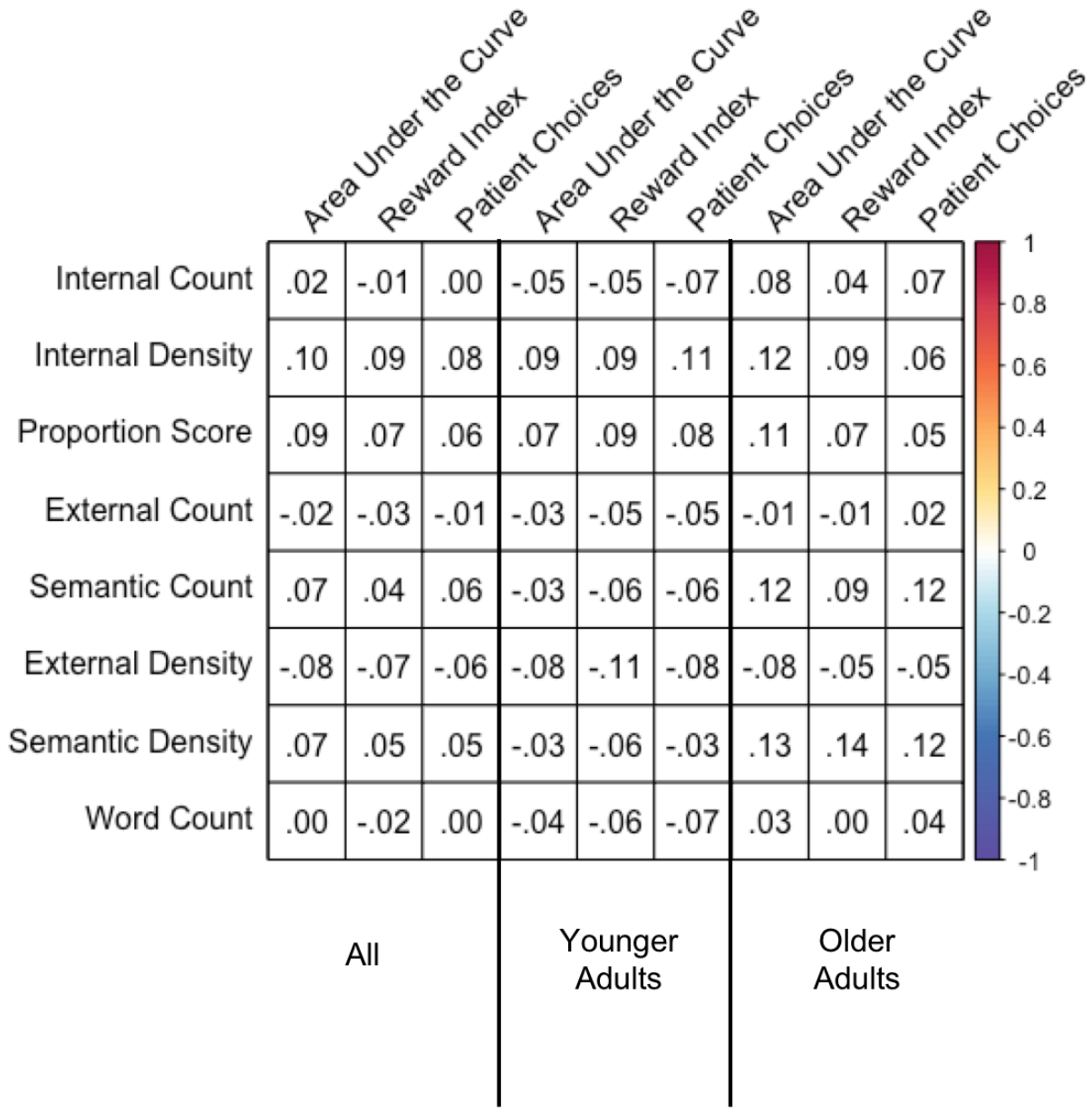
a




b




*Supplemental Figure 6.* Convergent Validity of the Autobiographical Interview with Episodic, Fluid, and Semantic Measures in Younger and Older Adults. Partial product-moment correlations controlling for gender between the 8 AI measures of interest and episodic, fluid, and semantic measures in a) younger adults and b) older adults. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction, ^ associations opposite to prediction ($p < .05$, uncorrected).




*Supplemental Figure 7.* Associations between the Autobiographical Interview with Measures of Temporal Discounting. Partial product-moment correlations between the 8 AI measures of interest and measures of temporal discounting across all participants controlling for age and gender, in younger adults controlling for gender, and in older adults controlling for gender. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate no significant association (p > .05, uncorrected).

a

| | IRI Empathic Concern | IRI Fantasy | IRI Personal Distress | IRI Perspective Taking | Mind in the Eye | Toronto Empathy |
|---|---|---|---|---|---|---|
| Internal Count | -.08 | .03 | -.14 | -.02 | .02 | -.02 |
| Internal Density | .04 | -.07 | -.05 | -.03 | .07 | -.03 |
| Proportion Score | .10 | .04 | .00 | -.07 | .20† | .03 |
| External Count | -.12 | -.01 | -.11 | -.03 | -.11 | -.06 |
| Semantic Count | -.10 | .03 | -.18 | .04 | -.07 | .01 |
| External Density | -.10 | -.08 | -.04 | .07 | -.25* | -.06 |
| Semantic Density | -.04 | .00 | -.14 | .14 | -.05 | .06 |
| Word Count | -.10 | .05 | -.11 | -.02 | -.01 | -.02 |

b

| | IRI Empathic Concern | IRI Fantasy | IRI Personal Distress | IRI Perspective Taking | Mind in the Eye | Toronto Empathy |
|---|---|---|---|---|---|---|
| Internal Count | .06 | .11 | .00 | .19† | .14 | .09 |
| Internal Density | -.10 | .13 | -.01 | -.07 | .06 | .01 |
| Proportion Score | -.12 | .18 | .04 | -.03 | .13 | .04 |
| External Count | .09 | -.05 | -.01 | .16 | .05 | .04 |
| Semantic Count | .15 | .06 | -.06 | .24* | .10 | .01 |
| External Density | .09 | -.20 | -.07 | -.01 | -.17 | .01 |
| Semantic Density | .21 | .07 | -.16 | .12 | .03 | -.02 |
| Word Count | .06 | .03 | .00 | .18 | .12 | .06 |

*Supplemental Figure 8.* Convergent Validity of the Autobiographical Interview with Measures of Social Cognition in Younger and Older Adults. IRI=Interpersonal Reactivity Index. Partial product-moment correlations between the 8 AI measures of interest and measures of personality controlling for gender within a) younger adults and b) older adults. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association, * unpredicted significant association with Bonferroni correction.

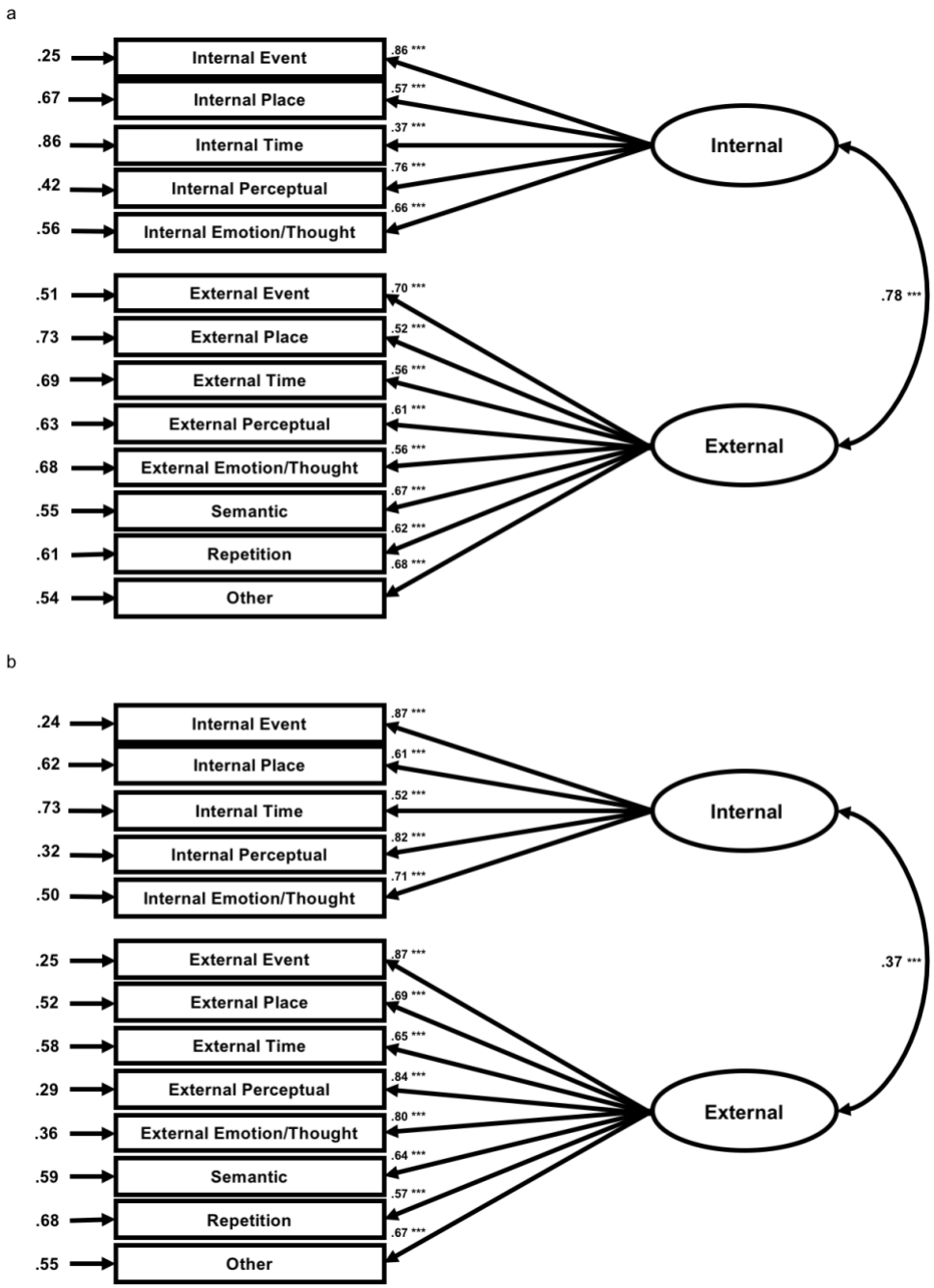
a

| | Openness/Intellect | Conscientiousness | Extraversion | Agreeableness | Neuroticism |
|---|---|---|---|---|---|
| Internal Count | .06 | .05 | .19 | -.04 | -.05 |
| Internal Density | -.04 | .05 | .01 | -.06 | -.13 |
| Proportion Score | -.06 | .01 | .02 | -.03 | -.10 |
| External Count | .06 | -.03 | .09 | -.04 | .08 |
| Semantic Count | .04 | .03 | .16 | .00 | -.02 |
| External Density | .05 | .01 | -.02 | .00 | .08 |
| Semantic Density | .09 | .06 | .10 | .05 | -.03 |
| Word Count | .06 | -.01 | .14 | -.02 | .02 |

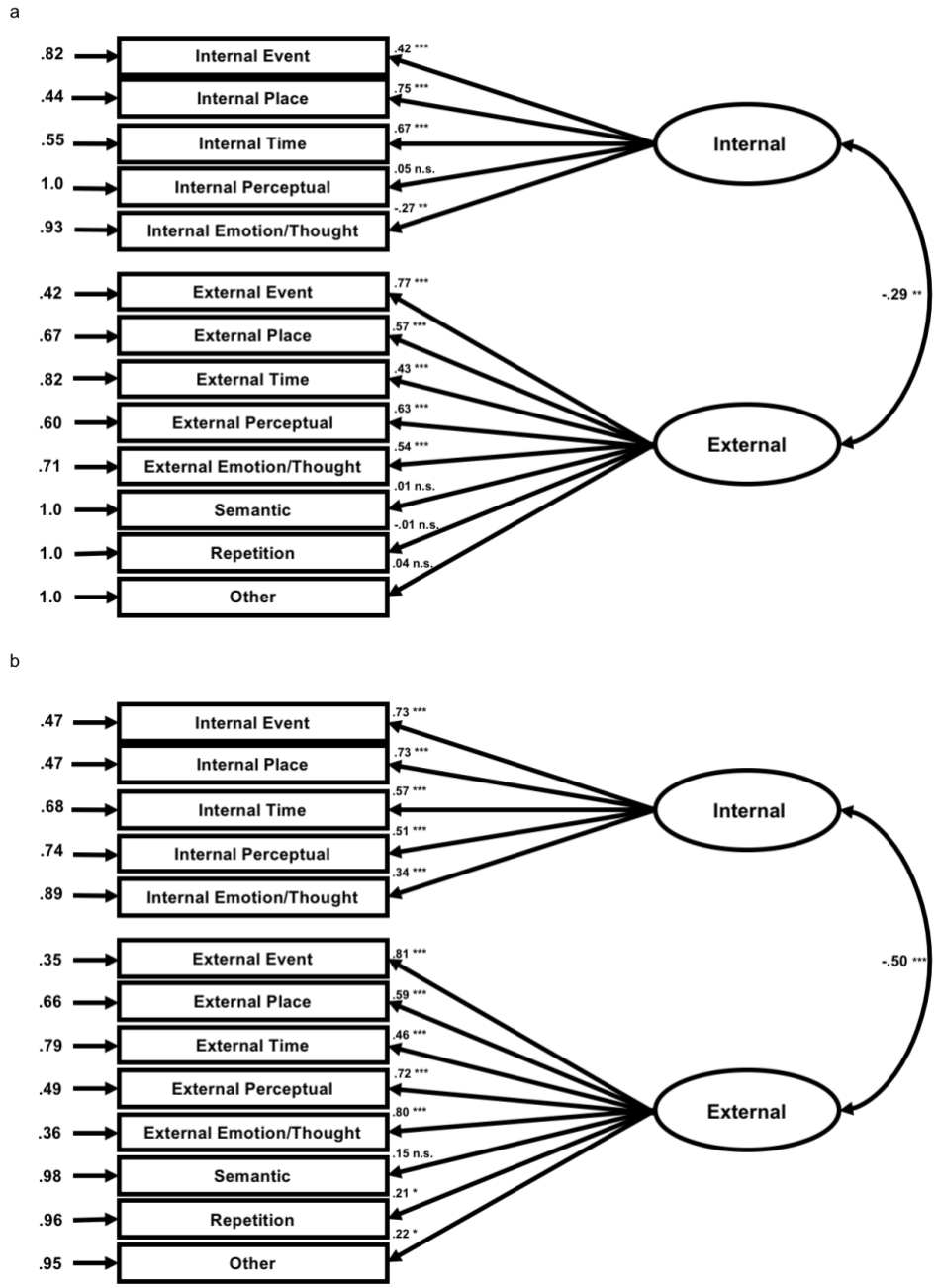
b

| | Openness/Intellect | Conscientiousness | Extraversion | Agreeableness | Neuroticism |
|---|---|---|---|---|---|
| Internal Count | .18† | .01 | -.04 | .03 | .05 |
| Internal Density | .10 | .07 | -.16 | -.08 | .15 |
| Proportion Score | .17† | .06 | -.18 | -.03 | .16 |
| External Count | -.02 | -.06 | .08 | .07 | -.12 |
| Semantic Count | .09 | -.03 | .07 | .05 | -.11 |
| External Density | -.19 | -.03 | .17 | .02 | -.14 |
| Semantic Density | .12 | .03 | .17 | .07 | -.09 |
| Word Count | .06 | -.03 | .03 | .05 | -.06 |

*Supplemental Figure 9.* Associations between the Autobiographical Interview with Measures of Personality in Younger and Older Adults. Partial product-moment correlations between the 8 AI measures of interest and measures of personality controlling for gender within a) younger adults and b) older adults. The matrix contains partial r values, which are color-coded by strength of the correlation. Cells with a white background indicate non-significant associations ($p > .05$, uncorrected). Color cells are significant ($p < .05$). † significant predicted association.




*Supplemental Figure 10.* Confirmatory Factory Analysis of Internal and External Detail Counts. The ellipses represent the latent variables of internal and external counts whereas the rectangles represent observed sub-categories of detail types which fall into each latent variable category, as represented by the straight arrows. Straight arrows between the latent and observed variables are shown with adjacent standardized regression coefficients (beta weights) and estimated with maximum likelihood estimation. The curved arrow represents the correlational relationship between the two latent variables and is shown with its adjacent correlation coefficient. Numbers at the end of the shorter arrows are standardized error variances. Diagrams are shown separately for a) younger adults and b) older adults, with groups analyzed in the same model. n.s. not significant; *p<.05; **p < .01; ***p<.001.




*Supplemental Figure 11.* Confirmatory Factory Analysis of Internal and External Density. n.s. not significant; Diagrams are shown separately for a) younger adults and b) older adults, with groups analyzed in the same model. n.s. not significant; *p<.05; **p < .01; ***p<.001.