BRIDGING NATURALISTIC AND LABORATORY ASSESSMENT OF MEMORY: THE BAYCREST MASK FIT TEST

Michael J. Armson[1,2], Hervé Abdi[3], and Brian Levine[1,2]

[1]Univ. of Toronto, Toronto, ON, Canada;

[2]Rotman Research Institute, Baycrest, Toronto, ON, Canada;

[3]The Univ. of Texas at Dallas, Dallas, TX, USA

Word Count: 4,329

Address correspondence to:
Michael Armson
Rotman Research Institute, Baycrest,
3560 Bathurst St,
Toronto, ON M6C 2M7
tel.: (416) 785-2500 X. 2826
Email: marmson@research.baycrest.org

Abstract

Autobiographical memory tests provide a naturalistic counterpoint to the artificiality of laboratory research methods, yet autobiographical events are uncontrolled and, in most cases, unverifiable. In this study, we capitalized on a scripted, complex naturalistic event—the Mask Fit Test (MFT), a standardized procedure required of hospital employees—to bridge the gap between naturalistic and laboratory memory assessment. We created a test of recognition memory for the MFT and administered it to 135 hospital employees who had undertaken the MFT at various points over the past five years. Multivariate analysis revealed two dimensions defined by accuracy and response bias. Accuracy scores showed the expected relationship to encoding-test delay, supporting the validity of this measure. Relative to younger adults, older adults' memory for this naturalistic event was better than would be predicted from the cognitive aging literature, a result consistent with the notion that older adults' memory performance is enhanced when stimuli are naturalistic and personally relevant. These results demonstrate that testing recognition memory for a scripted event is a viable method of studying autobiographical memory.

Key words: long-term memory; autobiographical memory; assessment; Multiple Correspondence Analysis (MCA); ecological validity

Introduction

In recent years, naturalistic memory tasks have been increasingly applied in experimental and applied clinical research. In particular, autobiographical memory tasks have been used in different experimental paradigms (e.g., neuroimaging paradigms, Cabeza & St. Jacques, 2007; Svoboda, McKinnon, & Levine, 2006) with healthy adults (e.g., Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002), but also with psychiatric (e.g., Williams et al., 2007) or neurological patients (e.g., Irish et al., 2011). Because of their self-relevance and time scales of operation, naturalistic tasks enable assessment of hypotheses that cannot be assessed with laboratory tasks, such as those concerning the involvement of the medial temporal lobe in remote memory (Bayley, Hopkins, & Squire, 2003; Rosenbaum et al., 2008; Sheldon & Levine, 2013) or the relationship between construction of past and future events (Addis, Wong, & Schacter, 2008; Sheldon & Levine, 2016). Autobiographical and laboratory tasks yield distinct patterns of brain activation (Gilboa, 2004; McDermott, Szpunar, & Christ, 2009; but see Rissman, Chow, Reggente, & Wagner, 2016). Moreover, laboratory test performance in individuals with highly superior or deficient naturalistic memory does not predict their everyday memory abilities (LePort et al., 2012; LePort, Stark, McGaugh, & Stark, 2016; Palombo, Alain, Söderlund, Khuu, & Levine, 2015).

Autobiographical memory tasks are therefore an important complement to laboratory tasks in the study of human memory, yet they have shortcomings relative to laboratory tasks. Autobiographical performance is usually assessed with free or cued recall, procedures that lack process-specificity. Group differences in naturalistic recall can be accounted by executive control processes (Dalgleish et al., 2007), and age-related reductions in episodic details on recall tasks generalize to unrelated tasks such as picture description (Gaesser, Sacchetti, Addis, & Schacter, 2011). Also, whereas the experimenter controls the encoding stimuli and context in laboratory tasks, by contrast, in most autobiographical memory tasks, participants select from their own life past events whose accuracy cannot be verified.

Vivid recollection of autobiographical memories can be evoked by highly specific retrieval cues from everyday life (Conway, Pleydell-Pearce, Whitecross, & Sharpe, 2003). With this in mind, some past studies have used prospective collection of autobiographical events in verbal diaries (Barclay & Wellman, 1986; Brewer, 1988; Conway, Collins, Gathercole, & Anderson, 1996; Levine et al., 2004; Linton, 1975; Thompson, 1982; Wagenaar, 1986) or photographs (Cabeza et al., 2004; St. Jacques, Conway, & Cabeza, 2011), to build a collection of retrieval cues for later testing. Although prospective collection of personal events allows for the assessment of accuracy of autobiographical memories, most of these previous studies contained heterogeneous and uncontrolled events. In one instance, when memory events were common among participants, the stimuli evoked contextual information for isolated locations rather than representing a fluid naturalistic event (Cabeza et al., 2004). While "flashbulb" memories of publicly shared emotional events may be less heterogeneous, they are subject to distortion through social interactions (Pezdek, 2003) and are of course impossible to control.

These limitations can be overcome by staged events that derive verified memoranda in a naturalistic context. Such paradigms have been used to study the effects of developmental conditions on real-life memory ability (Agnew & Powell, 2004; Bruck, London, Landa, & Goodman, 2007; Cooper, Vargha-Khadem, Gadian, & Maguire, 2011; Willoughby, McAndrews, & Rovet, 2014), the effects on memory accuracy and control of trauma and PTSD (McKinnon et al., 2015) and of semantic dementia (Adlam, Patterson, & Hodges, 2009), and the effects of retrieval manipulations such as reactivation (St. Jacques, Montgomery, & Schacter, 2014; St.

Jacques & Schacter, 2013). At the brain level, such paradigms can be used to dissociate retrieval of context (e.g., temporal or spatial; Fujii et al., 2004) and to relate memory accuracy to critical brain structures, such as the hippocampus (Willoughby et al., 2014) or the amygdala (Palombo, McKinnon, et al., 2015).

In the present study, we merged laboratory and naturalistic methods by assessing participants' memory for the Mask Fit Test (MFT), a scripted procedure mandated for hospital employees at Baycrest. The MFT is a standardized process where employees are fitted for a properly fitting respiratory mask in case of an outbreak at the hospital. We capitalized on the multisensory richness of this event to derive a large number of statements for testing in true/false recognition memory format, maximizing the control over assessment, increasing process-specificity, and allowing for the measurement of accuracy. To our knowledge, no study has attempted to assess episodic autobiographical memory of staged or verified events using a verbal recognition memory format (for examples with visual stimuli, see Adlam et al., 2009; Brewer & Treyens, 1981; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989; St. Jacques et al., 2014; St. Jacques & Schacter, 2013; see also Fujii et al., 2004 for neuroimaging example).

Our first goal was to establish the validity of this memory assessment procedure. As the MFT is required of all Baycrest employees, we were able to recruit a sample large enough (N = 135) to measure the latent structure of the MFT memory test (construct validity) and to evaluate predictions derived from well-replicated memory phenomena (criterion-related validity). Recognition responses were analyzed using Multiple Correspondence Analysis (MCA), a multivariate method tailored for nominal and ordinal data that identifies the dimensions (e.g., accuracy vs. response bias) that best characterize the variance in the data (Abdi & Valentin, 2007; Greenacre, 1984, 2007; Lebart, Morineau, & Warwick, 1984). Using the documented date of the MFT, we assessed the effects of remoteness at a temporal resolution on the order of months, well outside of the usual range for laboratory studies, but more recent than is typically the case in studies of autobiographical memory, with the prediction that memory accuracy would decline with time (Rubin & Schulkind, 1997).

As an ancillary goal, we assessed aging effects (within the limitation of our sample of non-retired Baycrest employees). To our knowledge, there is no aging study of recognition memory for verified autobiographical events (for exception from neuroimaging, see Maguire & Frith, 2003). Although aging is associated with reduced autobiographical memory specificity (Levine et al., 2002), aging effects on memory are reduced by the recognition test format (Craik & McDowd, 1987). Therefore, we did not have a strong hypothesis about aging effects on accuracy as measured by recognition memory.

Materials and methods

Participants

One hundred thirty-five participants [112 females, mean age = 35.70 years old (y.o.), *SD* = 12.13, mean education = 17.25 y.o., *SD* = 3.56] were recruited from the staff of Baycrest Hospital using intra-hospital advertisement through newsletters and public postings. Exclusion criteria included history of neurological disease, active significant medical illness, serious psychiatric illness, or substance abuse during the last 12 months. For analysis of recency effects, participants were divided into four age-, education-, and sex-matched groups according to time elapsed since MFT encoding: median 70 d (range: 29–130 d; *N* = 35), 249 d (range: 131–400 d; *N* = 35), 575 d (range: 401–700 d; *N* = 34), and 816 d (range: 701–1557 d; *N* = 31). For analysis of aging, the sample was divided into two age groups: < 45 y.o. and ≥ 45 y.o. This analysis was restricted to participants who were tested over one year following the MFT, as there was an insufficient number of older participants to assess aging effects in the more recent time periods. Thus the groups for this analysis consisted of 28 younger participants (22 females, mean age = 27.39 y.o., SD = 4.24) and 19 older participants (18 females, mean age = 54.10 y.o., SD = 5.09) matched for time elapsed since MFT encoding (younger: median 699 d; range: 393–1385 d; older: median 654 d; range: 384–1557 d).

Mask-fitting procedure at Baycrest Hospital

The MFT is a standardized, scripted procedure composed of various exercises to determine if an individual is wearing a properly fitting respiratory mask. As a government-mandated test, the MFT is a highly standardized and stereotyped procedure; everyone who takes it is exposed to the same elements. The exercises included spraying a bitter solution on each participant’s tongue to ensure taste sensitivity, donning a respiratory mask and a large white fume hood, spraying the bitter solution into the fume hood, reading a passage aloud, and making a series of head and body motions (see Appendix A for a synopsis of the MFT, as represented by our recognition test items).

Behavioural testing procedure

All 135 participants completed the MFT recognition test online. Participants also completed the Big Five Inventory (BFI; Benet-Martínez & John, 1998), the Object-Spatial Imagery Questionnaire (OSIQ; Blajenkova, Kozhevnikov, & Motes, 2006), the Survey of Autobiographical Memory (SAM; Palombo, Williams, Abdi, & Levine, 2013), and, in a subset of participants (*N* = 34) the Autobiographical Interview (Levine et al., 2002), although data from these instruments are not reported here. In exchange for completion of the online questionnaire, participants received a $10 gift card, as well as entry into a lottery draw for three prizes, one of two $50 gift cards or one $200 gift card. This study was approved by Baycrest’s ethics board.

Mask Fit Recognition Test

Memory for the MFT was assessed using true/false recognition memory. Because the Baycrest MFT is a standardized procedure, administered in a single location by the same technicians, it was possible to create items that genuinely reflected discrete elements of the MFT for all participants. Items were created with the assistance of the MFT technicians and refined through pilot testing. Participants rated 75 statements, presented one at a time on the computer screen (see Appendix A). This procedure provided 38 true statements and 37 lures. After reading each statement, participants rated the confidence of their recognition on a six-point Likert scale (1 = “certain false,” 2 = “unsure false,” 3 = “very unsure false”, 4 = “very unsure true,” 5 = “unsure true,” 6 = “certain true”). Baycrest employees are required to repeat the MFT at regular intervals (approximately every two years). Seventy-eight of our participants had undergone the MFT once, 21 twice, and 36 three or more times. Participants who had completed more than one MFT were instructed to base their recognition memory responses on their most recent MFT exposure.

Statistical analyses

MFT recognition memory data were analyzed using MCA from the package R ExPosition (Abdi & Valentin, 2007; Abdi & Béra, 2014; Greenacre, 1984, 2007; Lebart et al., 1984). MCA is more suitable than traditional data reduction techniques (e.g., principal components analysis) for data involving ordinal scales, such as Likert scales, where the response categories correspond to an ordinal rather than interval or ratio scales. Data reduction techniques such as MCA hold an advantage over univariate analyses because they investigate for patterns of correlation among variables and therefore avoid the “multiple comparison problem” that originates when performing multiples statistical tests on large sets of variables. In addition, because MCA transforms qualitative data (i.e., collected on nominal ordinal scales), into quantitative data, it allows for more meaningful further analysis than simply analyzing the raw ordinal data according to univariate methods. MCA represents data by computing a series of optimal and pairwise orthogonal (linear) combinations—called factors, dimensions, or components—of the original variables. In MCA, components are computed for both observations (i.e., participants) and variables (i.e., test items); the value of a component for an observation or a variable is called a factor score. These factor scores can be used as coordinates to represent observations and variables as points on maps. In these maps, similar observations (or variables) are near each other, so relationships between observations and between the variables can be visually identified. Each dimension extracts a portion of the total variance (called inertia) of the data. The variance explained by a dimension, is called the eigenvalue of this dimension (note that, in MCA, this eigenvalue is always smaller than 1).

We used the "doubling of rating" coding scheme (also called "fuzzy-coding," see Greenacre, 2007, 2014). Doubling represents each question by two poles: one representing "certain false" and the other representing "certain true." The response for one question is represented by two variables, where the values measure the proximity of the response to the poles, with the constraint that the sum of these two numbers is equal to one. For example, on a scale going from 1 to 6, the answer 1 is represented by the values [1, 0], a value of 2 would be coded as [4/5, 1/5], a value of 4 would be coded [2/5, 3/5], a value of 6 would be coded [0, 1]. With this method of coding, a question is represented on factor maps by a line connecting the point representing the smallest value (i.e., "certain false") to the point representing the largest value ("certain true"). Note that these lines always pass by the origin of the space. With the doubled coding scheme used, the cosine of the angle between two lines approximates the correlation between the two questions represented by the lines (i.e., a small angle indicates a large correlation and two orthogonal lines indicate a zero correlation). The length of a line approximates the variance of the question represented by the line. When one extremity of the line is near the origin, this indicates that this extremity was often used by the respondents (and, conversely, an extremity far away from the origin indicates that it was rarely used).

To assess if the overall analysis was reliable, we computed the total inertia (a statistic equivalent to the independence χ^2 statistic) and tested its significance using a permutation test in which participants' responses were permuted independently for each question. A new analysis was performed for which the total inertia was computed as well as the eigenvalue associated to each dimension. This procedure was repeated 1,000 times. The proportion of observed inertia and eigenvalues larger than the observed values provides an empirical *p* value. In all reported analyses, the total inertia of the analysis was larger than all the permuted values (i.e., $p < .001$). In order to identify the reliable components, we used the same permutation test.

For each dimension, the stability of the factor scores of the modalities of the variables was assessed with a bootstrapping procedure by which 1,000 new data sets were created by resampling with replacement from the original data set. Each data set was projected as supplementary elements on the original solution (a procedure sometimes called "partial bootstrap," see Abdi & Williams, 2010; Abdi & Béra, 2014, for details on this procedure) in order to obtain bootstrapped factor scores for the responses. The ratios of the means of these factor scores to their standard deviations give bootstrap ratios that can be interpreted like *t*-statistics. Responses with a bootstrap ratio of 3 or greater for a given dimension (roughly corresponding to a *p* value of .05 after Bonferroni correction for the total number of questions) were retained for further analysis and interpretation.

Results

The first seven dimensions of the MCA were significant at the $p < .05$ significance level. However, the scree of the eigenvalues suggested that the first two dimensions (Dimension 1: $\lambda = 0.041$, $p = .001$, $\tau = 10\%$; Dimension 2: $\lambda = 0.033$, $p = .001$, $\tau = 6\%$, where λ is the eigenvalue and τ is the percent of inertia extracted by a given dimension) stood out from the remainder. From this point on, these two dimensions were used as the axes of our graphs onto which we plotted the original MFT recognition data, as well as projecting group means and confidence intervals (and this is how the two dimensions are represented in the figures presented here).

The first step in our graphical analyses was to plot the MFT recognition test items onto the factor space created by the two retained dimensions (see Figure 1). In this plot, each MFT item was coded according to its item number, its category (blue = target; red = lure), and its response (T = “True”; F = “False”). Because incorrect responses (misses and false alarms) were clustered on the left side of Dimension 1—while correct responses (hits and correct rejections) were clustered on the right side of Dimension 1—we interpreted Dimension 1 as reflecting overall memory accuracy. The left side of Dimension 1 represents inaccurate recognition memory, while the right side represents accurate memory. For Dimension 2, responses of “true” to either a true or false item were clustered toward the top, while responses of “false” to either item type were clustered toward the bottom. We therefore interpreted Dimension 2 as representing response bias. Considering the extremity of the lines, it can be seen that correct responses (hits and correct rejections) were endorsed with a high degree of confidence (short lines, see methods) whereas incorrect responses (misses and false alarms) were endorsed with lower confidence (long lines).

We next projected our participants onto the factor space. Each participant was plotted by using his or her factor scores for Dimensions 1 and 2 as x and y coordinates. On the plot, participants were colour-coded according to the four recency groups (see Figure 2). The mean responses of each group were ordered sequentially along Dimension 1 according to recency, a pattern indicating a positive association between recency and MFT recognition accuracy, with significant differences between the most recently (70 d) and most remotely (816 d) tested groups, as well as between the 70 d and the 249 d group and between the 249 d and 816 d groups. The two middle groups showed nearly complete overlap, whereas the two most remote groups showed only slight overlap. There was no significant effect of recency on mean Dimension 2 scores, a pattern indicating that response bias was not affected by recency. The above pattern of results was confirmed by a 4x2 factorial ANOVA comparing the four recency groups in terms of their Dimension 1 and 2 scores from the MCA. This ANOVA showed a significant main effect of recency group on Dimension 1 [$F(3, 131) = 27.711$, $p < .001$] but not on Dimension 2 scores [$F(3, 131) = 1.792$, $p = .152$], with post hoc (Tukey HSD) tests indicating significant differences between the 70 d group and all other groups, and between the 816 d group and the 249 d and 575 d groups.

As mentioned above, some participants experienced the MFT multiple times, which could potentially confound effects of recency. However, there was no significant relationship between repetitions of the MFT and scores on Dimension 1 [$r(133) = .013$, $p = .877$] or Dimension 2 [$r(133) = -.132$, $p = .126$]. Similarly, there were no significant differences between participants who experienced the MFT just once versus more than once on Dimension 1 [$t(133) = 0.143$, $p = .886$] or Dimension 2 [$t(133) = 0.710$, $p = .479$] scores.

To assess the effect of age on our MCA factor scores, we compared Dimension 1 and 2 scores for those above or below 45 years of age tested at least one year post MFT (see Methods). Our age groups differed significantly in terms of repetitions of the MFT, with older participants having experienced the MFT more times than younger participants [younger: $M = 1.464$; $SD = .693$; older: $M = 2.737$; $SD = 1.284$; $t(25.160) = 3.947$, $p = .001$]. Thus, to control for the potential effect of repetitions of the MFT, we entered number of MFTs as a covariate in our analysis. Younger and older groups had similar scores for both Dimension 1 (younger: $M = -.079$; $SD = .154$; older: $M = -.115$; $SD = .205$; $F(1, 44) = 0.49$, $p = .825$] and Dimension 2 [younger: $M = .065$; $SD = .174$; older: $M = -.044$; $SD = .201$; $F(1, 44) = 0.206$, $p = .652$].

Discussion

Although cognitive science studies of memory are intended to foster understanding of memory as it operates in day-to-day life, naturalistic and laboratory methods rarely coincide. In the present study, we leveraged the standardization of a common hospital procedure to apply laboratory-based recognition methods of assessing memory to a naturalistic event. We established the construct and criterion-related validity of this new platform with multivariate methods through analysis of the recency effect. Although a small number of studies have assessed recognition pictures depicting staged events, verbal responses are more typical of naturalistic memory operations (e.g., answering questions or talking about events). From a practical perspective, verbal assessment is preferred where image quality is difficult to control (e.g., Internet testing) or impossible (e.g., Sheldon, Amaral, & Levine, 2016).

The inverse relationship between MFT recency and recognition memory test performance supports the validity of the MFT recognition memory test. Moreover, the nature of the mask fitting event as a mildly arousing and precisely dated event enabled probing of memory at a critical period for memory consolidation (when extrapolated from animal research, Anagnostaras, Maren, & Fanselow, 1999; Bontempi, Laurent-Demir, Destrade, & Jaffard, 1999; Winocur, Moscovitch, Caruana, & Binns, 2005; see Sheldon & Levine, 2013). This period, on the scale of months (Rubin, 1982; Rubin & Wenzel, 1996), is typically opaque to laboratory methods—where memory is assessed at hours or days with relatively mundane stimuli such as word lists—as well as to retrospective autobiographical memory assessment methods—where highly personally significant events are probed years or decades after their occurrence.

Within autobiographical memory, older adults show reduced richness and increased production of non-episodic information (Levine et al., 2002). Older adults are also more susceptible to memory distortion following misleading information compared to young adults (Wylie et al., 2014). Similarly, on naturalistic tasks testing memory for perceived and imagined events, older adults show a deficit on recall compared to young adults (Hashtroudi, Johnson, & Chrosniak, 1990; Hashtroudi, Johnson, Vnek, & Ferguson, 1994). Despite these previous findings, we found no effect of age on accuracy in recognizing details from the MFT one year later (for similar finding, see St. Jacques, et al., 2014). Recognition can be robust to aging effects when it is supported by non-episodic processes such as familiarity (Craik & McDowd, 1987; Jennings & Jacoby, 1993; Stark, Yassa, Lacy, & Stark, 2013). Additionally, as mentioned earlier, older adults' performance on real-life memory tasks is often better than expected relative to laboratory memory tasks (Aberle & Rendell, 2010; Rendell & Craik, 2000). Older adults also perform at a similar or higher level than young adults on real life tasks where information is meaningful, self-relevant, or pleasant (Mather & Carstensen, 2005; Salthouse, 2004). It is acknowledged that our ability to detect aging effects may have been affected by the relatively young age of our sample, which was limited by the necessity to test Baycrest employees.

While the majority of participants had undergone the MFT just once, some had experienced it multiple times, which could have affected recognition memory performance. That said, there was no significant relationship between multiple MFT exposures and Dimension 1 and 2 scores. Given that the interval between MFT exposures is at least two years (according to Baycrest policy), previous experiences with the MFT may have been relatively inaccessible to consciousness at the time of testing (Rubin, 1982; Rubin & Wenzel, 1996). It is also possible that multiple MFT exposures could have rendered the details of the MFT more resistant to forgetting (Moscovitch, Cabeza, Winocur, & Nadel, 2016) and created a stable general event representation (Renoult, Davidson, Palombo, Moscovitch, & Levine, 2012). In any case, the presence or absence of prior MFT exposure does not affect our main conclusions concerning the validity of our recognition memory testing procedure.

The assessment of memory for naturalistic events is of primary relevance to memory researchers for both practical and theoretical reasons. Laboratory assessment methods are assumed to pertain to memory in real-life—an assumption that extends to standardized clinical tests used for diagnosis of memory impairment. Although laboratory and naturalistic memory assessments provide non-overlapping results, the heterogeneity of naturalistic events impedes their assessment. Using a standardized, scripted naturalistic event, we derived valid measures of memory functioning through analysis of verbal recognition memory performance, which is more controlled than free recall. As noted in the introduction, few studies have used staged events for assessment of episodic autobiographical memory, and this is the first such study to use verbal recognition memory (as opposed to photographs or free recall) as a testing platform that can be accommodated to other staged or naturalistic events. This study was observational, in that we did not construct the tested event, which would enable more control over event characteristics and study-test delay for specific hypothesis testing (e.g., St. Jacques et al., 2014; St. Jacques & Schacter, 2013, Willoughby et al., 2014). Future studies should compare free recall and recognition for both laboratory and naturalistic stimuli (Diamond & Levine, 2014), which would allow for more precise characterization of memory accuracy under different retrieval contexts.

Acknowledgments

This research was supported by grants to BL from the Canadian Institutes of Health Research (MOP 62963 and the National Institutes of Health (RO1 MH076067). HA would like to acknowledge the support of an EURIAS fellowship at the Paris Institute for Advanced Studies (France), with the support of the European Union's 7th Framework Program for research, and from funding from the French State managed by the "Agence Nationale de la Recherche (program: Investissements d'avenir, ANR-11-LABX-0027-01 Labex RFIEA+)". We would also like to thank members of the Occupational Health and Safety (OH&S) Department at Baycrest who contributed to this project. Specifically, former OH&S Director Margaret Cernigoj assisted with development of the project in its early stages, and former MFT administrators Robert Ferraro and Gaynor Leigh assisted with development of the questionnaire items

References

Abdi, H., & Béra, M. (2014). Correspondence analysis. In R. Alhajj and J. Rokne (Eds.), *Encyclopedia of Social Networks and Mining.* New York: Springer Verlag. pp 275–28. doi:10.1007/978-1-4614-6170-8_140

Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. *Encyclopedia of Measurement and Statistics*, 651–657. Retrieved from http://wwwpub.utdallas.edu/~herve/Abdi-MCA2007-pretty.pdf

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*, 433–459. doi:10.1002/wics.101

Aberle, I., & Rendell, P. (2010). The age prospective memory paradox: young adults may not give their best outside of the lab. *Developmental Psychology*, *46*, 1444–1453. doi:10.1037/a0020718

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychological Science, 19*, 33–41. doi: 10.1111/j.1467-9280.2008.02043.x

Adlam, A. L., Patterson, K., & Hodges, J. R. (2009). “I remember it as if it were yesterday”: Memory for recent events in patients with semantic dementia. *Neuropsychologia*, *47*, 1344–1351. doi:10.1016/j.neuropsychologia.2009.01.029

Agnew, S. E., & Powell, M. B. (2004). The effect of intellectual disability on children's recall of an event across different question types. *Law and Human Behavior*, *28*, 273. doi:10.1023/B:LAHU.0000029139.38127.61

Anagnostaras, S. G., Maren, S., & Fanselow, M. S. (1999). Temporally graded retrograde amnesia of contextual fear after hippocampal damage in rats: within-subjects examination. *The Journal of neuroscience*, *19*, 1106–1114. Retrieved from http://www.jneurosci.org/content/19/3/1106.full

Barclay, C. R., & Wellman, H. M. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory and Language*, *25*(1), 93–103. doi:10.1016/0749-596X(86)90023-9

Bayley, P. J., Hopkins, R. O., & Squire, L. R. (2003). Successful recollection of remote autobiographical memories by amnesic patients with medial temporal lobe lesions. *Neuron*, *38*, 135–144. doi:10.1016/S0896-6273(03)00156-9

Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, *75*, 729–50. doi:10.1037/0022-3514.75.3.729

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: a new self-report imagery questionnaire. *Applied Cognitive Psychology*, *20*, 239–263. doi:10.1002/acp.1182

Bontempi, B., Laurent-Demir, C., Destrade, C., & Jaffard, R. (1999). Time-dependent reorganization of brain circuitry underlying long-term memory storage. *Nature*, *400*, 671–675. doi:10.1038/23270

Brewer, W. F. (1988). Memory for randomly sampled autobiographical events. In U. Neisser & E. Winograd (Eds.), Remembering reconsidered: Ecological and traditional approaches to the study of memory. Emory symposia in cognition, 2 (pp. 21–90). New York: Cambridge University Press. doi:10.1017/CBO9780511664014.004

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive psychology, 13*, 207–230. doi:10.1016/0010-0285(81)90008-6

Bruck, M., London, K., Landa, R., & Goodman, J. (2007). Autobiographical memory and suggestibility in children with autism spectrum disorder. *Development and Psychopathology*, *19*, 73–95. doi:10.1017/S0954579407070058

Cabeza, R., Prince, S. E., Daselaar, S. M., Greenberg, D. L., Budde, M., Dolcos, F., ... & Rubin, D. C. (2004). Brain activity during episodic retrieval of autobiographical and laboratory events: An fMRI study using a novel photo paradigm. *Journal of cognitive neuroscience*, *16*, 1583–1594. doi:10.1162/0898929042568578

Cabeza, R., & St. Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, *11*, 219–227. doi:10.1016/j.tics.2007.02.005

Cohen, G., Conway, M. A., & Maylor, E. A. (1994). Flashbulb memories in older adults. *Psychology and Aging*, *9*, 454. doi:10.1037/0882-7974.9.3.454

Conway, M. A., Collins, A. F., Gathercole, S. E., & Anderson, S. J. (1996). Recollections of true and false autobiographical memories. *Journal of Experimental Psychology: General*, *125*, 69–95. doi:10.1037/0096-3445.125.1.69

Conway, M. A., Pleydell-Pearce, C. W., Whitecross, S. E., & Sharpe, H. (2003). Neurophysiological correlates of memory for experienced and imagined events. *Neuropsychologia*, *41*, 334–340. doi:10.1016/S0028-3932(02)00165-3

Cooper, J. M., Vargha-Khadem, F., Gadian, D. G., & Maguire, E. A. (2011). The effect of hippocampal damage in children on recalling the past and imagining new experiences. *Neuropsychologia*, *49*, 1843–1850. doi:10.1016/j.neuropsychologia.2011.03.008

Craik, F. I., & Mcdowd, J. M. (1987). Age Differences in Recall and Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 474–479. doi:10.1037/0278-7393.13.3.474

Dalgleish, T., Williams, J. M. G., Golden, A. M. J., Perkins, N., Barrett, L. F., Barnard, P. J., ... & Watkins, E. (2007). Reduced specificity of autobiographical memory and depression: the role of executive control. *Journal of Experimental Psychology: General*, *136*, 23–42. doi:10.1037/0096-3445.136.1.23

Davidson, P. S., & Glisky, E. L. (2002). Is flashbulb memory a special instance of source memory? Evidence from older adults. *Memory*, *10*(2), 99–111. doi: 10.1080/09658210143000227

Diamond, N., & Levine, B. (2014, November). Naturalistic and laboratory encoding contexts dissociate subjective and objective measures of episodic memory in older adults. Poster presented at the annual meeting of the Society for Neuroscience, Washington, DC, USA. Abstract retrieved from http://www.abstractsonline.com/Plan/ViewAbstract.aspx?sKey=713dccc3-8bff-482d-803b-cede93b276c1&cKey=de0d4d3d-2dc1-4f07-a73b-bdec12cdc1b5&mKey=%7b54C85D94-6D69-4B09-AFAA-502C0E680CA7%7d

Fujii, T., Suzuki, M., Okuda, J., Ohtake, H., Tanji, K., Yamaguchi, K., … Yamadori, A. (2004). Neural correlates of context memory with real-world events. *NeuroImage*, *21*, 1596–1603. doi:10.1016/j.neuroimage.2004.01.005

Gilboa, A. (2004). Autobiographical and episodic memory—one and the same?: Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia*, *42*, 1336–1349. doi:10.1016/j.neuropsychologia.2004.02.014

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*, 80–84. doi:10.1037/a0021054

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. New York: Academic Press.

Greenacre, M. (2007). *Correspondence Analysis in Practice.* Boca Raton: CRC Press.

Greenacre, M. (2014). Data doubling and fuzzy coding. In J. Blasius & M. Greenacre (Eds), *Visualization and Verbalization of Data*. (pp 239–270). Boca Raton: CRC Press. Retrieved from https://books.google.ca/books?hl=en&lr=&id=0hQyAwAAQBAJ&oi=fnd&pg=PA239&dq=Data+doubling+and+fuzzy+coding&ots=8QcoxDL1Ts&sig=Lp7T36142I3j7g8UNFtCB4lDcwo&redir_esc=y#v=onepage&q=Data%20doubling%20and%20fuzzy%20coding&f=false

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging*, *5*, 119–126. doi:10.1037/0882-7974.5.1.119

Hashtroudi, S., Johnson, M. K., Vnek, N., & Ferguson, S. A. (1994). Aging and the effects of affective and factual focus on source monitoring and recall. *Psychology and Aging*, *9*, 160–170. doi:10.1037/0882-7974.9.1.160

Irish, M., Hornberger, M., Lah, S., Miller, L., Pengas, G., Nestor, P. J., ... & Piguet, O. (2011). Profiles of recent autobiographical memory retrieval in semantic dementia, behavioural-variant frontotemporal dementia, and Alzheimer's disease. *Neuropsychologia*, *49*, 2694–2702. doi:10.1016/j.neuropsychologia.2011.05.017

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging*, *8*, 283–293. doi:10.1037//0882-7974.8.2.283

Lebart, L., Morineau, A., & Warwick, K. (1984) *Multivariate Descriptive Statistical Analysis*. London: Wiley.

LePort, A., Mattfeld, A., Dickinson-Ansom, H., Fallon, J., Stark, C., Kruggel, F., … McGaugh, J. (2012). Behavioral and neuroanatomical investigation of highly superior autobiographical memory (HSAM). *Neurobiology of Learning and Memory*, *98*, 78–92. doi:10.1016/j.nlm.2012.05.002

LePort, A. K., Stark, S. M., McGaugh, J. L., & Stark, C. E. (2016). A cognitive assessment of highly superior autobiographical memory. *Memory*. Advance online publication. doi:10.1080/09658211.2016.1160126

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*, 677–689. doi:10.1037//0882-7974.17.4.677

Levine, B., Turner, G. R., Tisserand, D., Hevenor, S. J., Graham, S., & McIntosh, A. R. (2004). The functional neuroanatomy of episodic and semantic autobiographical remembering: a prospective functional MRI study. *Cognitive Neuroscience, Journal of*, *16*, 1633–1646. doi:10.1162/0898929042568587

Linton, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 376–404). San Francisco: Freeman & Company.

Maguire, E. A., & Frith, C. D. (2003). Aging affects the engagement of the hippocampus during autobiographical memory retrieval. *Brain*, *126*, 1511–1523. doi:10.1093/brain/awg157

Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: the positivity effect in attention and memory. *Trends in Cognitive Sciences*, *9*, 496–502. doi:10.1016/j.tics.2005.08.005

McDermott, K. B., Szpunar, K. K., & Christ, S. E. (2009). Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, *47*, 2290–8. doi:10.1016/j.neuropsychologia.2008.12.025

McKinnon, M. C., Palombo, D. J., Nazarov, A., Kumar, N., Khuu, W., & Levine, B. (2015). Threat of Death and Autobiographical Memory A Study of Passengers From Flight AT236. *Clinical Psychological Science*, *3*, 487–502. doi:10.1177/2167702614542280

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic Memory and Beyond: The Hippocampus and Neocortex in Transformation. *Annual Review of Psychology*, *67*, 105–134. doi:10.1146/annurev-psych-113011-143733

Palombo, D. J., Alain, C., Söderlund, H., Khuu, W., & Levine, B. (2015). Severely deficient autobiographical memory (SDAM) in healthy adults: A new mnemonic syndrome. *Neuropsychologia*, *72*, 105–118. doi:10.1016/j.neuropsychologia.2015.04.012

Palombo, D. J., McKinnon, M. C., McIntosh, A. R., Anderson, A. K., Todd, R. M., & Levine, B. (2015). The Neural Correlates of Memory for a Life-Threatening Event: An fMRI Study of Passengers From Flight AT236. *Clinical Psychological Science*, *4*, 312–319. doi:10.1177/2167702615589308

Palombo, D. J., Williams, L. J., Abdi, H., & Levine, B. (2013). The survey of autobiographical memory (SAM): a novel measure of trait mnemonics in everyday life. *Cortex*, *49*, 1526–40. doi:10.1016/j.cortex.2012.08.023

Pezdek, K. (2003). Event memory and autobiographical memory for the events of September 11, 2001. *Applied Cognitive Psychology, 17*, 1033–1045. doi:10.1037/0278-7393.15.4.587

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 587. doi:10.1037/0278-7393.15.4.587

Rendell, P., & Craik, F. (2000). Virtual week and actual week: Age-related differences in prospective memory. *Applied Cognitive Psychology*, *62*, 43–62. doi:10.1002/acp.770

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

Rissman, J., Chow, T. E., Reggente, N., & Wagner, A. D. (2016). Decoding fMRI Signatures of Real-world Autobiographical Memory Retrieval. *Journal of cognitive neuroscience*, *28*, 604–620. doi:10.1162/jocn_a_00920

Rosenbaum, R. S., Moscovitch, M., Foster, J. K., Schnyer, D. M., Gao, F., Kovacevic, N., ... & Levine, B. (2008). Patterns of autobiographical memory loss in medial-temporal lobe amnesic patients. *Journal of Cognitive Neuroscience*, *20*, 1490–1506. doi:10.1162/jocn.2008.20105

Rubin, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 21–38. doi:10.1016/S0022-5371(82)90423-6

Rubin, D. C., & Schulkind, M. D. (1997). The distribution of autobiographical memories across the lifespan. *Memory & Cognition*, *25*, 859–866. doi:10.3758/BF03211330

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological review*, *103,* 734. doi:10.1037/0033-295X.103.4.734

Salthouse, T. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*, 140–144. doi: 10.1111/j.0963-7214.2004.00293.x

Sheldon, S., Amaral, R., & Levine, B. (2016). Individual differences in visual imagery predict event memory. *Memory*. Advance online publication. doi:10.1080/09658211.2016.1178777

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. doi:10.1016/j.neuroimage.2013.06.082

Sheldon, S., & Levine, B. (2016). The role of the hippocampus in memory and mental construction. *Annals of the New York Academy of Sciences*, *1369,* 76-92. doi: 10.1111/nyas.13006

St. Jacques, P. L., Conway, M. A., & Cabeza, R. (2011). Gender differences in autobiographical memory for everyday events: retrieval elicited by SenseCam images versus verbal cues. *Memory*, *19*, 723–732. doi:10.1080/09658211.2010.516266

St. Jacques, P. L., Montgomery, D., & Schacter, D. L. (2014). Modifying memory for a museum tour in older adults: Reactivation-related updating that enhances and distorts memory is reduced in ageing. *Memory*, *23*, 876–887. doi:10.1080/09658211.2014.933241

St. Jacques, P. L., & Schacter, D. L. (2013). Modifying Memory Selectively Enhancing and Updating Personal Memories for a Museum Tour by Reactivating Them. *Psychological Science*, *24*, 537–543. doi:10.1177/0956797612457377

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*, 2442–9. doi:10.1016/j.neuropsychologia.2012.12.014

Svoboda, E., McKinnon, M., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia*, *44*, 2189–2208. doi:10.1016/j.neuropsychologia.2006.05.023

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*, 455–461. doi: 10.1111/1467-9280.02453

Thompson, C. P. (1982). Memory for unique personal events: The roommate study. *Memory & Cognition*, *10*, 324–332. doi:10.3758/BF03202424

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive psychology*, *18*, 225–252. doi:10.1016/0010-0285(86)90013-7

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*, 122–148. doi:10.1037/0033-2909.133.1.122

Willoughby, K. A., McAndrews, M. P., & Rovet, J. F. (2014). Accuracy of episodic autobiographical memory in children with early thyroid hormone deficiency using a staged event. *Developmental Cognitive Neuroscience*, *9*, 1–11. doi:10.1016/j.dcn.2013.12.005

Winocur, G., Moscovitch, M., Caruana, D. A., & Binns, M. A. (2005). Retrograde amnesia in rats with lesions to the hippocampus on a test of spatial memory. *Neuropsychologia*, *43*, 1580–1590. doi:10.1016/j.neuropsychologia.2005.01.013

Wylie, L. E., Patihis, L., McCuller, L. L., Davis, D., Brank, E., Loftus, E. F., & Bornstein, B. (2014). Misinformation Effect in Older versus Younger Adults: A Meta-Analysis and Review. In M. P. Toglia, D. F. Ross, J. Pozzulo, & E. Pica (Eds.), *The elderly eyewitness in court*. UK: Psychology Press. doi:10.4324/9781315813936

Appendix

MFT Recognition Test Items

#	Item	True/False (T/F)
Q1	The mask fit test took place in the Posluns Building.	T
Q2	The mask fit test took place on the third floor.	T
Q3	The hallway outside the test room had white walls.	T
Q4	The hallway had a painting of swans swimming in a pond.	F
Q5	There was a table in the hallway.	T
Q6	There were clipboards on the table.	T
Q7	You practiced reading a passage that you would later read while wearing the mask.	F
Q8	There were magazines to read while waiting for the test to begin.	F
Q9	You were asked to wash your hands before the test.	F
Q10	You filled out the form in the hallway before entering the test room.	T
Q11	The form was printed on light blue paper.	F
Q12	The form was two pages long.	F
Q13	The form asked if you had asthma.	T
Q14	The form asked your consent to have your test results shared as part of a national mask fit database.	F
Q15	As you walked in the test room, there were chairs lined up against the right-hand wall.	F
Q16	The chairs had a silver metal frame with wooden arm rests.	T
Q17	There was a couch up against one wall.	F
Q18	The room had white walls.	T
Q19	The room had a dark carpet.	T
Q20	There was a window in the door to the test room.	F
Q21	There was an abstract painting in the test room.	F
Q22	There was a coat rack in the test room.	F
Q23	There was a large filing cabinet in the test room.	T
Q24	There was a chart on the wall of different types of respirator masks.	F
Q25	There was a large receptacle to hold respirator masks.	T
Q26	There was a water dispenser in the test room.	T
Q27	There was a coffee machine in the test room.	F
Q28	The test administrator spoke with an accent.	T
Q29	The test administrator had dark brown hair.	T
Q30	The test administrator wore a white lab coat.	F
Q31	The test administrator wore glasses.	F
Q32	The Sensitivity Test occurred before you tried on a mask.	T
Q33	You were told to open your mouth when the solution was sprayed.	T

Q34	The test administrator sprayed the solution multiple times in quick succession.	F
Q35	You were standing up when the solution was sprayed.	F
Q36	You were told to close your eyes when the solution was sprayed.	F
Q37	The solution that was sprayed tasted sour.	F
Q38	The test administrator told you not to swallow immediately after the solution was sprayed.	F
Q39	You were offered a drink of water after the solution was sprayed.	T
Q40	You briefly held masks up to your face to eliminate ones that obviously did not fit.	F
Q41	All masks were light blue with elastic head bands.	F
Q42	You watched the test administrator put on the mask as an example.	T
Q43	You practiced putting on and taking off the mask several times in a row.	F
Q44	The test administrator adjusted the straps of your mask for you.	F
Q45	A mirror was used to assess the fit of the mask.	F
Q46	You read a passage aloud before you put on the fume hood.	F
Q47	The passage was presented on laminated paper.	T
Q48	The passage listed each different colour in a rainbow.	F
Q49	The passage discussed the phenomenon of a "double rainbow".	F
Q50	The passage talked about the legendary pot of gold at one end of the rainbow.	T
Q51	The passage talked about Judy Garland's song "Over the Rainbow".	F
Q52	You were standing up while reading the passage.	T
Q53	You put on the fume hood before the solution was sprayed again.	T
Q54	The test administrator took the fume hood from the top of a table.	T
Q55	The test administrator took the fume hood from your left.	T
Q56	The majority of the surface of the fume hood was white.	T
Q57	The hood had a square of clear plastic to see through.	T
Q58	You had to pull on draw strings to make the hood fit.	F
Q59	The solution was sprayed through a hole in the fume hood.	T
Q60	You were standing up for this spray test.	T
Q61	The head movement exercises happened directly after a spray test.	T
Q62	The head movement exercises included moving your head side to side.	T
Q63	The head movement exercises included moving your head up and down.	T
Q64	One exercise involved turning around in a circle.	F
Q65	One exercise involved bending forward at the waist.	T
Q66	One exercise involved making a frown while wearing the mask.	F
Q67	The final spray test happened before you took off the fume hood.	T
Q68	You were standing up during the final spray test.	T

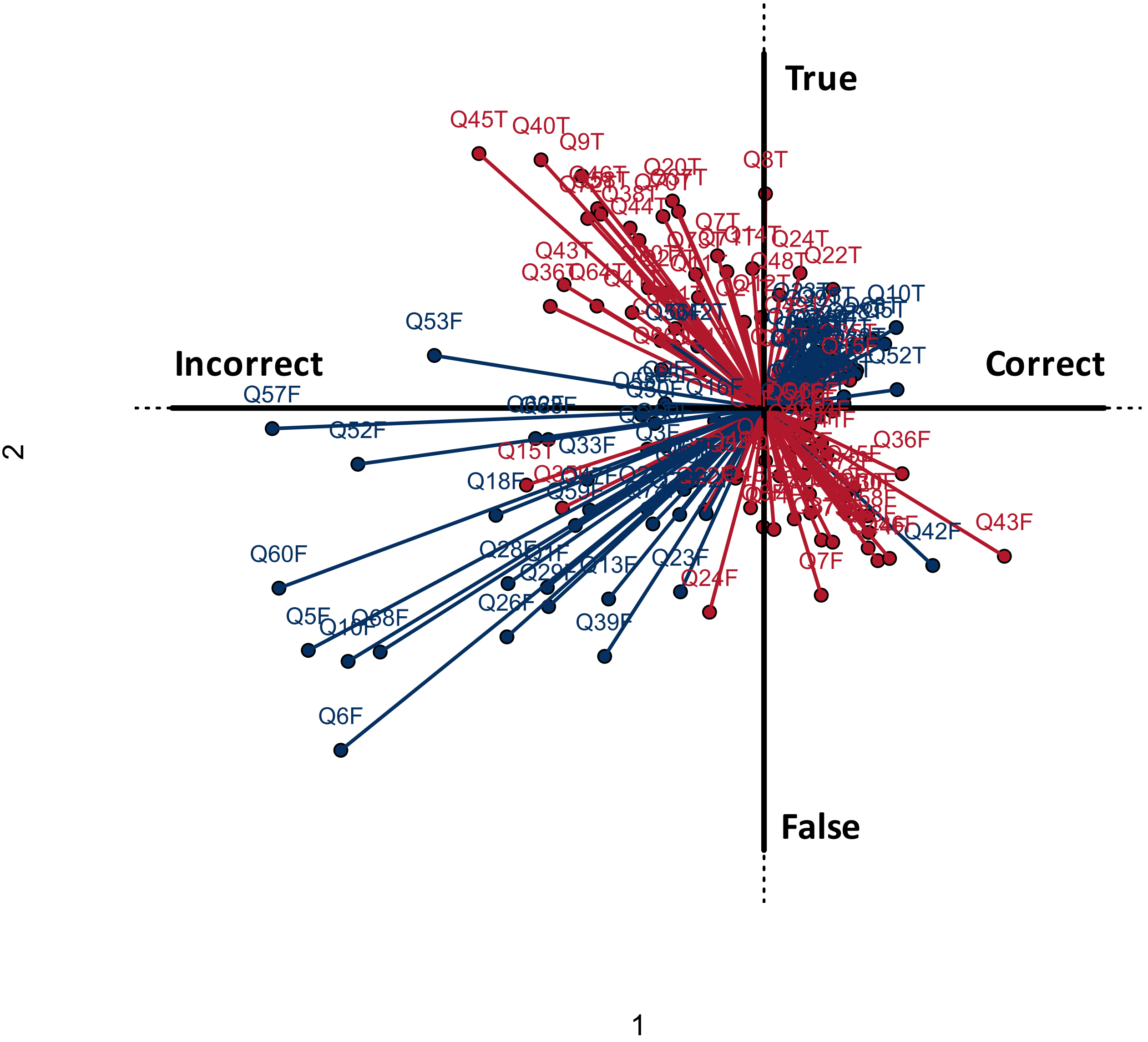
Q69	You were instructed to remove the mask by pulling it away from your chin.	T
Q70	The mask that fit you properly had a plastic circular disk attached to the front.	F
Q71	You had to fill out an evaluation form before leaving.	F
Q72	You were given a pamphlet describing the safe use of respirator masks.	F
Q73	The bowl of candy was on the opposite side of the room from the door.	F
Q74	There was hand sanitizer next to the candy bowl.	T
Q75	The candy bowl was made of plastic.	T

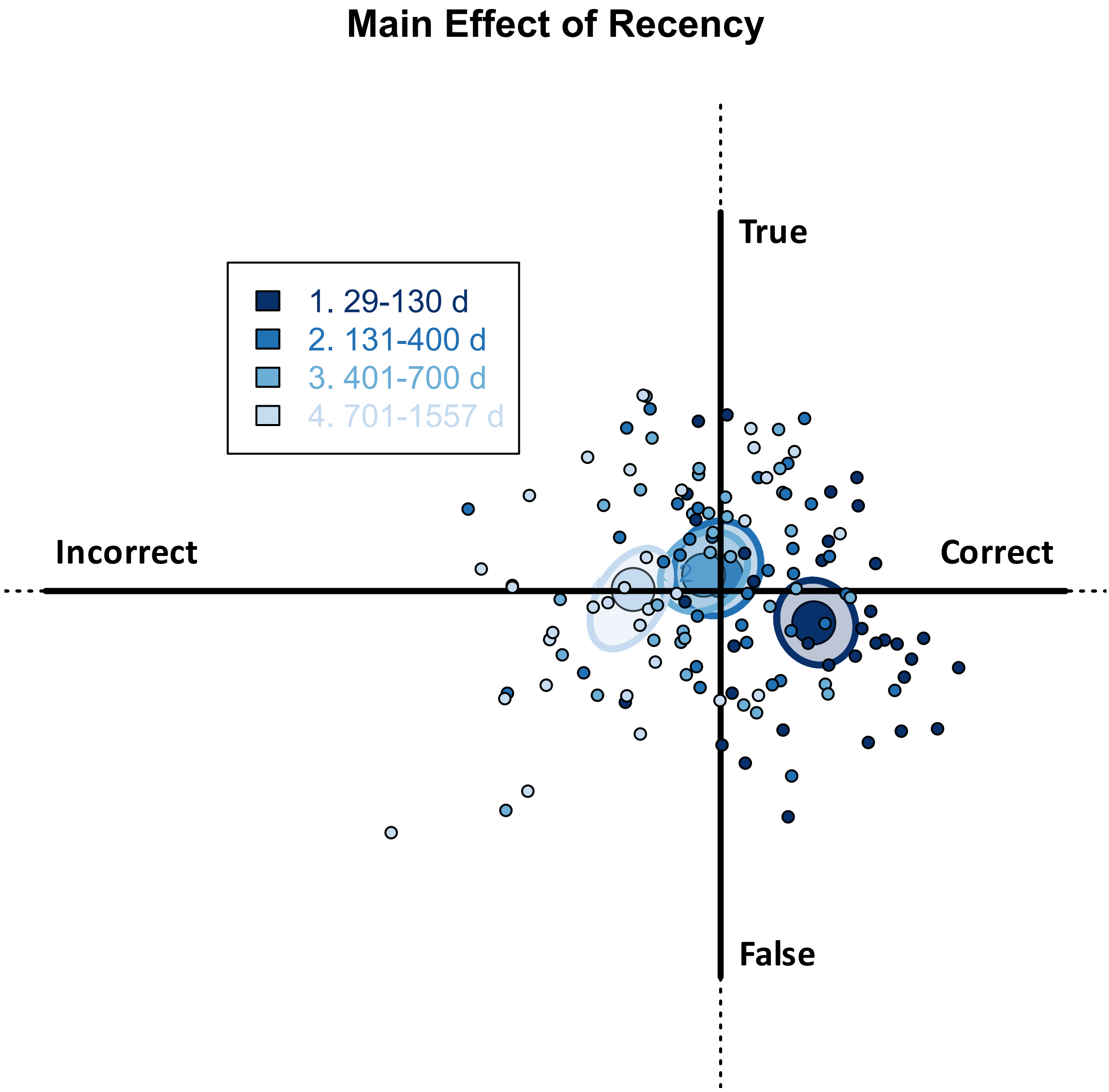
Figure Captions

Figure 1. Plot of MFT recognition test items within the factor space defined by Dimensions 1 and 2. In the plot, each item is coded according to its item number, its category (blue = target; red = lure), and its response (T = "True"; F = "False"). Dimension labels were interpreted based on the pattern of item responses loading on that dimension. Dimension 1, reflecting overall accuracy, is defined by incorrect responses (misses and false alarms; left) and correct responses (hits and correct rejections; right). Dimension 2, reflecting response bias, is defined by "true" responses at the top and "false" responses at the bottom.

Figure 2. Effect of Recency. Individual participants are projected into the factor space, with each dot representing a participant colour-coded for the recency of their MFT prior to study participation. Recency groups are represented by large circles whose x and y coordinates reflect their mean factor scores. The ellipses surrounding the large circles represent the variance of each recency group as bootstrap-derived 95% confidence intervals.

MFT Recognition Test Items





Main Effect of Recency
True
False
Incorrect
Correct
1. 29-130 d
2. 131-400 d
3. 401-700 d
4. 701-1557 d