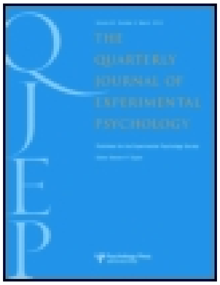

The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

Individual differences in forced-choice recognition memory: Partitioning contributions of recollection and familiarity

Ellen M. Migo[a], Joel R. Quamme[b], Selina Holmes[cd], Andrew Bendell[c], Kenneth A. Norman[e], Andrew R. Mayes[c] & Daniela Montaldi[c]

[a] King's College London, Department of Psychological Medicine, Institute of Psychiatry, London, UK

[b] Psychology Department, Grand Valley State University, Allendale, MI, USA

[c] School of Psychological Sciences, University of Manchester, Manchester, UK

[d] Shropshire Enablement Team, Shropshire Rehabilitation Centre, Shrewsbury, UK

[e] Psychology Department, Princeton University, Princeton NJ, USA

Published online: 06 May 2014.

To cite this article: Ellen M. Migo, Joel R. Quamme, Selina Holmes, Andrew Bendell, Kenneth A. Norman, Andrew R. Mayes & Daniela Montaldi (2014): Individual differences in forced-choice recognition memory: Partitioning contributions of recollection and familiarity, The Quarterly Journal of Experimental Psychology, DOI: 10.1080/17470218.2014.910240

To link to this article: http://dx.doi.org/10.1080/17470218.2014.910240

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2014
http://dx.doi.org/10.1080/17470218.2014.910240

Individual differences in forced-choice recognition memory: Partitioning contributions of recollection and familiarity

Ellen M. Migo[1], Joel R. Quamme[2], Selina Holmes[3,4], Andrew Bendell[3], Kenneth A. Norman[5], Andrew R. Mayes[3], and Daniela Montaldi[3]

[1]King's College London, Department of Psychological Medicine, Institute of Psychiatry, London, UK
[2]Psychology Department, Grand Valley State University, Allendale, MI, USA
[3]School of Psychological Sciences, University of Manchester, Manchester, UK
[4]Shropshire Enablement Team, Shropshire Rehabilitation Centre, Shrewsbury, UK
[5]Psychology Department, Princeton University, Princeton NJ, USA

In forced-choice recognition memory, two different testing formats are possible under conditions of high target–foil similarity: Each target can be presented alongside foils similar to itself (forced-choice corresponding; FCC), or alongside foils similar to other targets (forced-choice noncorresponding; FCNC). Recent behavioural and neuropsychological studies suggest that FCC performance can be supported by familiarity whereas FCNC performance is supported primarily by recollection. In this paper, we corroborate this finding from an individual differences perspective. A group of older adults were given a test of FCC and FCNC recognition for object pictures, as well as standardized tests of recall, recognition, and IQ. Recall measures were found to predict FCNC, but not FCC performance, consistent with a critical role for recollection in FCNC only. After the common influence of recall was removed, standardized tests of recognition predicted FCC, but not FCNC performance. This is consistent with a contribution of only familiarity in FCC. Simulations show that a two-process model, where familiarity and recollection make separate contributions to recognition, is 10 times more likely to give these results than a single-process model. This evidence highlights the importance of recognition memory test design when examining the involvement of recollection and familiarity.

Keywords: Recognition memory; Recall; Recollection; Familiarity; Memory.

Recognition memory is our ability to decide whether something has been encountered before. Previous work suggests that recognition judgements are supported by two cognitive experiences; recollection and familiarity (Mandler, 1980). Familiarity is a feeling of memory of varying strength, whereas recollection involves retrieval of contextual associations from a previous encounter. For example, recollection might involve remembering what you were thinking when something was

Correspondence should be addressed to Ellen Migo, Academic Unit of Neuropsychiatry, King's College London, Department of Psychological Medicine, Institute of Psychiatry, Adamson Centre, South Wing Block D, St Thomas's Hospital, Westminster Bridge Road, London SE1 7EH, UK. E-mail: ellen.migo@kcl.ac.uk

E. Migo and J. Quamme contributed equally to this work.
We would like to thank Daniel Thayanandan for his help with data collection and Becky Haynes and Lara Harris for their suggestions on a previous version of this manuscript.
Part of this work was supported by National Institutes of Health (NIH) [grant number R01 MH069456-01].

seen previously. In contrast, familiarity is a feeling of memory strength produced by the reencounter, without remembering any such associations. In this study we used an individual differences approach to explore whether recollection and familiarity are differentially associated with performance on two forms of forced-choice recognition in a group of healthy older adults.

Prior work suggests that the contribution of recollection and familiarity to recognition performance depends on test format under conditions of high target–foil similarity. For instance, when recognition is tested in a yes/no (YN) format ("did you study this item, yes or no?"), studies show that participants' ability to distinguish previously studied target words (e.g., "cats") from their nonstudied switched-plurality forms (e.g., "cat") depends on recollection (Hintzman, Curran, & Oppy, 1992; Rotello, Macmillan, & Van Tassel, 2000; see Migo, Montaldi, Norman, Quamme, & Mayes, 2009, for evidence using visual stimuli). In forced-choice formats, where a target item is shown alongside new items (i.e., "which of these did you study?"), the contribution of recollection and familiarity to performance may depend on the manner in which target items are paired with foil items.

In a forced-choice corresponding (FCC) format, target items are shown alongside foils similar to themselves on a test trial (e.g., "did you study 'cat' or 'cats'?"). Neuropsychological studies suggest that FCC performance is supported at least in part by familiarity, whereas YN performance requires recollection. Holdstock et al. (2002) examined the recognition performance of patient Y.R., who showed an isolated recollection deficit alongside preserved familiarity. Y.R. was significantly impaired on a YN test for object pictures but performed at normal levels on a four-alternative FCC test (Holdstock et al., 2002). This dissociation has been replicated in two groups of amnestic mild cognitive impairment (aMCI) patients (Westerberg et al., 2013; Westerberg et al., 2006), a group for whom there is growing evidence of impaired recollection and preserved familiarity, at least for visual stimuli (Anderson et al., 2008; Belleville, Ménard, & Lepage, 2011; Deason, Hussey, Budson, & Ally, 2012; Embree, Budson, & Ally, 2012; Hudon, Belleville, & Gauthier, 2009; O'Connor & Ally, 2010; Schefter et al., 2013; Serra et al., 2010).

Forced-choice recognition may also be tested in a "noncorresponding" format (FCNC), in which target items are shown alongside foils similar to other target items (Tulving, 1981). Here, participants might study "cats" and "dog", then be asked to decide between "cats" and "dogs" at test. Migo et al. (2009) directly compared FCC and FCNC performance in young healthy participants using the Holdstock et al. (2002) object pictures. FCNC performance was selectively reduced relative to FCC when participants responded based only on feelings of familiarity. This suggests that like YN, FCNC performance relies on recollection, whereas FCC performance can be supported by familiarity assessments. This combined evidence from different patient groups and healthy controls suggests that these test formats are differentially dependent on recollection and familiarity.

This pattern (familiarity supports FCC, but not YN or FCNC) is predicted by the complementary learning systems (CLS) model (Norman & O'Reilly, 2003). The CLS is a neurocomputational model of hippocampal and medial temporal function that assumes that hippocampus and perirhinal cortex are responsible for recollection and familiarity processing, respectively (for reviews of dual-process theories of medial temporal lobe function see Brown & Aggleton, 2001; Eichenbaum, Yonelinas, & Ranganath, 2007; Montaldi & Mayes, 2010; Norman, 2010). In tests with high target–foil similarity, the perirhinal/familiarity component performs poorly on YN and FCNC tests because the variation in memory strength amongst targets is large relative to the differences between targets and their similar foils. Consequently, false alarms occur because a foil to a strongly encoded target may feel more familiar than that to a poorly encoded target. In an FCC test, however, the small differences in familiarity between the targets and their corresponding foils are highly reliable across trials, supporting successful familiarity discriminations. The hippocampal component can use recall of specific details to both reject the highly similar foils and to accept targets in any test format. The CLS model, therefore, makes the clear prediction that FCC performance may be supported by familiarity assessments, whereas FCNC and YN performance should require recollection (see Migo et al., 2009, for further discussion of this prediction).

In the present study, we used an individual differences approach to find converging evidence for the differential contribution of recollection and familiarity to FCC and FCNC. A combined test of FCC and FCNC recognition for object pictures with high target–foil similarity was administered to a group of healthy older adults, alongside standardized neuropsychological tests assessing recall and recognition memory, as well as a brief measure of IQ. Healthy older adults were used to ensure a reasonable degree of variance in standardized test performance among participants, in order to facilitate the analyses.

We examined how FCC and FCNC recognition relate to standardized measures of recall and recognition across individuals, in a manner consistent with separate, and differential, contributions of recollection and familiarity. Specifically, we used correlation, regression, and latent-factor simulations to explore whether FCC and FCNC exhibit different variance partitions attributable to familiarity and recognition. We derived theoretical expectations concerning the role of recollection and familiarity-related variance in FCC, FCNC, and standardized recall and recognition. Following prior individual differences work on sources of variance underlying recall and recognition (Quamme, Yonelinas, Widaman, Kroll, & Sauve, 2004; Unsworth & Brewer, 2009), we assumed that shared variance with standardized recall measures would be attributable in part to recollection, but not to familiarity, whereas shared variance with standardized recognition could be attributable to both recollection and familiarity.

To test for a recollection-related variance partition in FCC and FCNC, we examined the extent to which each of these measures was predicted by standardized recall. To test for a familiarity-related variance partition, we examined the degree to which FCC and FCNC could be predicted by standardized recognition after variance shared by recall was removed. Following CLS predictions (Norman & O'Reilly, 2003) and prior empirical work on FCC and FCNC (Migo et al., 2009), we expected to find a significant recollection-related variance partition in FCNC and a significant familiarity-related variance partition in FCC, but not FCNC. Note, we did not have strong a priori predictions about the contribution of recollection to FCC (in principle, there is no reason that subjects could not use recollection to make FCC judgements). We also conducted latent variable simulations to determine whether the results were uniquely consistent with separate contributions of recollection and familiarity, or whether alternative models (such as a single memory process) would be just as likely to generate similar variance partitions. This approach allowed us to directly test between theories of recognition memory, where there is continued debate over whether a single- or dual-process model is more appropriate (Eichenbaum, Sauvage, Fortin, Komorowski, & Lipton, 2012; Montaldi & Mayes, 2010; Ranganath, 2010; Wais, 2013).

Method

Participants

A total of 68 participants (mean age 71 years, range 50–85, 28 male) took part in the study. Ethical approval was obtained from the School of Psychological Sciences Research Ethics Committee, University of Manchester and the Psychiatry, Nursing & Midwifery Research Ethics Subcommittee, King's College London.

Neuropsychological memory test battery

The standardized memory tests used to estimate memory performance were the Wechsler Memory Scale–Third Edition Abbreviated (WMS–IIIA; Logical Memory and Family Pictures subtests; Wechsler, 1998) and the Doors and People Test (DP; Baddeley, Emslie, & Nimmo-Smith, 1994). We also used the Wechsler Test of Adult Reading (WTAR; Wechsler, 2001) to predict full-scale IQ (predicted FSIQ).

The Logical Memory subtest of WMS–IIIA is a story recall measure. Participants are read two short stories and are then asked to repeat each back immediately and then again after a delay of 25–35 min. One story is read once, and another is read out twice. In the Family Pictures subtest of WMS–IIIA, participants are shown four pictures for 10 seconds each, where each picture presents members of a family in a scene. Immediately after seeing all four pictures, then again after a delay of 25–35 min, participants are asked to describe who was in each picture, giving details of their location and of their activity. WMS–IIIA, therefore, gives measures of immediate and delayed recall, which are also combined to give an index of total recall. These composite scores are age scaled with a mean of 100 and a standard deviation of 15, analogous to IQ.

DP has four component tests: recognition of doors, recall of people, recall of shapes, and recognition of names. The doors test is a visual recognition test where participants study pictures of doors and are asked to pick the studied door from a choice of four alternative pictures. The test has two parts, where Part A is easier since Part B uses more similar doors. Each test consists of 12 trials. The names test is a similar four-alternative forced-choice recognition test where stimuli are names. As with the doors test, there are two parts of 12 items, where the second set of items uses more similar foils. The people test is a recall measure of first name–surname combinations, and the shapes test is a recall measure of two-dimensional simple abstract shapes. Scaled scores are calculated for each component and are then combined to give recall (composed of the people and shapes tests) and recognition (composed of the doors and names tests) age-scaled scores. DP therefore gives independent measures of recognition and recall, based on different stimuli. Scores are reported as age-scaled scores with a mean of 10 and standard deviation of 3, as calculated by the test norms.

Stimuli

Digital greyscale photos of similar alternatives of everyday items were taken to create 27 sets of between 15 and 32 pictures. Picture similarity information was obtained using a stimulus sorting procedure (Goldstone, 1994), where participants moved pictures on a computer screen so that more similar pictures were closer together. The distances across picture sets were standardized, and backgrounds were manually removed. Further work on standardizing the similarity of these pictures, where the full set of stimuli are provided, has also been published (Migo, Montaldi, & Mayes, 2013).

Twenty-four picture quartets were selected with one target with three equally and highly similar foils (see Figure 1 for examples). The similarity level was higher than DP Part B (Baddeley et al., 1994; see Migo, Weiss, Norman, Mayes, & Montaldi, 2008, for similarity comparison), which was designed to have a relatively high level of target–foil similarity. For comparison purposes the similarity levels used in this experiment correspond to a similarity value of approximately 2,400 from the rescaled tables in Migo et al. (2013). Eight additional picture quartets for the training task were selected without similarity measures, but were judged to have a level of target–foil similarity comparable to that for critical test items.

Experimental test procedure

Participants completed a practice test before the main experiment. This ensured that all participants realized how similar the targets and foils would be and meant that they had seen both types of test trial before the experiment began. The practice test consisted of eight studied items and six test trials (three FCC, three FCNC). The procedure was based on the use of similar tests with patients (see Figure 2). In the main experiment, participants were shown 24 pictures twice each, one at a time, for three seconds each, in a fully randomized order. On the first presentation they made a natural versus man-made judgement. Verbal responses were recorded, and the intertrial interval was two seconds. On the second presentation they were asked try to remember as much about them as possible. A one-minute delay was filled with arithmetic questions.

In the test phase, participants were shown 24 quartets of pictures: 12 FCC trials and 12 FCNC

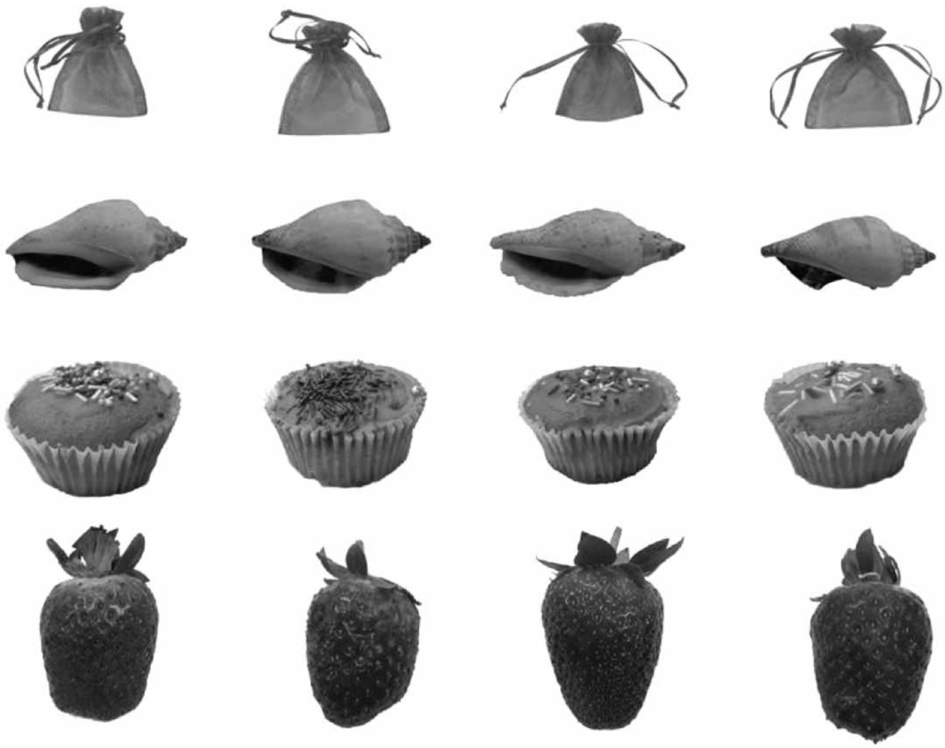

Figure 1. *Example stimuli used in the experiment. In all examples, the target picture is on the left, with its three similar foils alongside.*

trials (Figure 2). Each picture and its similar foils appeared in only one format for each participant. The test trials were presented in an ABBA design between formats (six A, 12 B, six A). The allocation of pictures to format and the starting format of the test phase were counterbalanced across participants. The spatial position of the correct picture was counterbalanced, and the order of pictures within a format was fully randomized. Participants were asked to decide which picture had been seen before, without time pressure, and their verbal responses were recorded. The maximum possible score for each test format was therefore 12, with absolute scores converted to proportions.

Results

Summary scores from all participants are presented in Table 1. Statistical analysis was carried out using SPSS Version 20, and all tests report two-tailed significance results. Effect sizes were calculated using G*Power 3.1.3 (Faul, Erdfelder, Lang, & Buchner, 2007).

Performance on the FCC and FCNC formats was above chance [FCC: $t(67) = 19.974$, $p < .001$, Cohen's $d = 2.42$; FCNC: $t(67) = 14.309$, $p < .001$, Cohen's $d = 1.74$] and significantly better in the FCC format [paired t test: $t(67) = 2.644$, $p = .010$, Cohen's $d = 0.32$], without floor or ceiling effects. An analysis of variance (ANOVA) comparing test format (FCC versus FCNC), test order (FCC first versus FCNC first), and picture list allocation (List 1 FCC versus List 1 FCNC) only showed a significant effect of test format [$F(1, 64) = 6.661$, $p = .012$; next largest $F = 2.846$]. Alpha measures of internal consistency were calculated for each picture list in each test format, giving values of .71 (List 1) and .51 (List 2) for FCC and .68

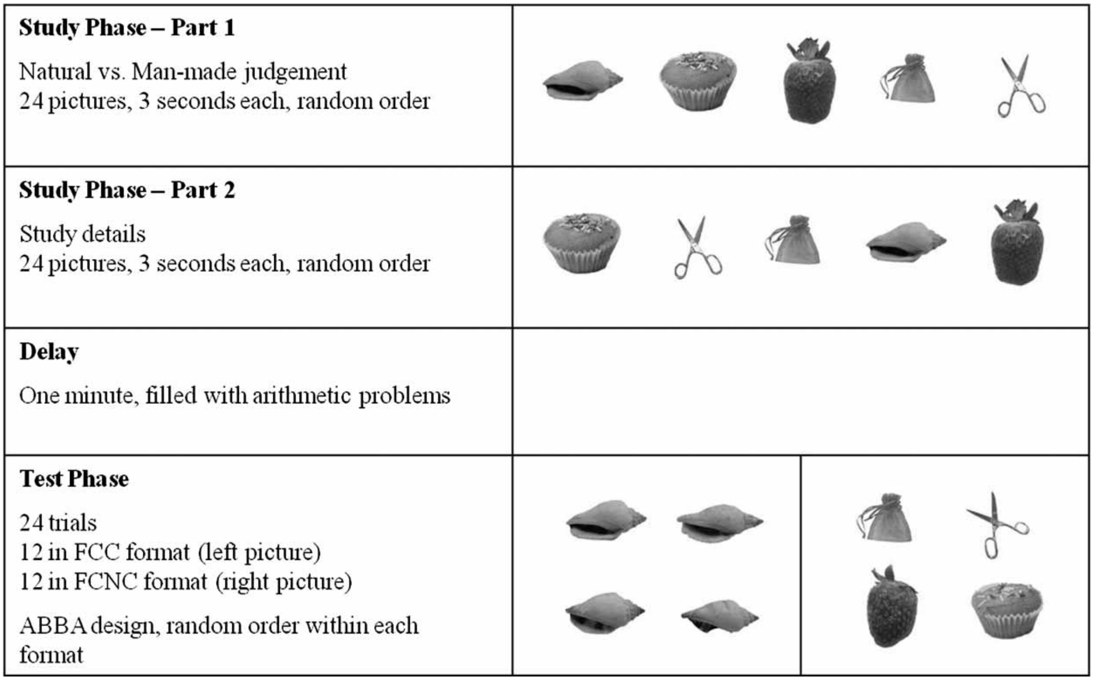

Figure 2. *Procedure summary. FCC = forced-choice corresponding; FCNC = forced-choice noncorresponding. See text for full details.*

(List 1) and .67 (List 2) for FCNC. These values are broadly in line with reliabilities from other experimental tests (e.g., Unsworth & Brewer, 2009), especially considering the relatively short test length of 12 items.

Three types of analysis were carried out: an exploratory correlation analysis, a theoretically motivated hierarchical regression analysis, and finally a simulation to directly compare the plausibility of single- and dual-process explanations for the results.

Table 1. *Mean performance on all tests*

Test		*Performance*	*Range*	*Skew*	*Kurtosis*
DP	Recall	12.44 (3.44)	5–18	−0.12	−1.06
	Recognition	13.16 (2.95)	6–18	−0.20	−0.61
WMS–IIIA	Immediate	108.09 (13.43)	74–136	−0.04	−0.28
	Delay	110.74 (13.67)	76–137	−0.15	−0.22
	Total	108.84 (13.40)	81–134	−0.05	−0.50
Predicted FSIQ		110.29 (5.98)	92–118	−1.04	0.90
Experimental tests	FCC	.71 (.19)	.25–1.00	−0.80	−0.06
	FCNC	.63 (.22)	.00–1.00	−0.25	−0.38

Note: DP = Doors and People Test; WMS–IIIA= Wechsler Memory Scale–Third Edition Abbreviated; FSIQ = full-scale intelligence quotient; FCC = forced-choice corresponding; FCNC = forced-choice noncorresponding. Where appropriate (DP, WMS–IIIA, predicted FSIQ), these are age-adjusted scores. All values to two decimal places. For DP recall this averages measures from the People and Shapes subtests, and for DP recognition this averages measures from the Doors and Names subtests. Immediate WMS–IIIA recall reliability averages immediate Logical Memory (LM) and Family Pictures (FP) measures, whereas delayed WMS–IIIA recall reliability averages delayed LM and FP measures. Standard deviations in parentheses.

Correlation results

The correlations between FCC, FCNC, and the standardized tests are shown in Table 2. FCNC performance correlated significantly with all measures of recall, recognition, and IQ, whereas FCC performance correlated significantly with recognition but only with one of the recall scores (WMS–III delayed recall; $r = .24$, $p = .047$). Thus, as expected, recall performance was more associated with FCNC, and recognition performance was associated with both FCC and FCNC. Age was only significantly (negatively) correlated with FCC performance and (positively) with predicted IQ. Age correlations with the standardized tests are not expected as these are age-scaled scores.

Hierarchical regression results

To separate the common influences of recollection and familiarity on standardized recall and recognition measures on one hand and experimental forced-choice measures on the other hand, two 2-stage hierarchical regressions were carried out. The recall measures may predict forced-choice performance because of a shared dependence on recollection, whereas recognition may predict forced-choice performance on the basis of shared recollection, shared familiarity, or both. Therefore, it should be possible to first isolate any influence of recollection by predicting FCC and FCNC from only recall and then subsequently to isolate any influence of shared familiarity by adding recognition as a predictor in a second step. We used DP recall and recognition as predictor variables, as well as total memory from WMS–IIIA. The three measures from WMS–IIIA correlated very highly with each other (see Table 2), and therefore only one was included to avoid multicollinearity. Total memory was selected since it gives a performance indicator incorporating both immediate and delayed recall.

The recall measures were entered in Stage 1 of the regression to examine variance explained by contribution of recollection common to both recall and forced-choice measures. If R^2 is significantly greater than zero at Stage 1, this suggests that a significant proportion of variance in forced-choice recognition is attributable to recollection. In the second stage, the DP recognition scores were added as a third predictor to examine whether additional variance is explained by the common contribution of familiarity, given that recollection has been already accounted for. If there is a significant increase in R^2 at Stage 2, this suggests that a significant proportion of variance in forced-choice recognition is attributable to familiarity.

The results from the analysis are shown in Table 3. For FCNC, recall tests accounted for a significant proportion of the variance in Stage 1 [$R^2 = .154$, $F(2, 65) = 5.901$, $p = .004$]. However, the addition of the recognition test in Stage 2 did not significantly improve the variance explained by the model [$\Delta R^2 = .038$, $F(1, 64) = 3.007$, $p = .088$], suggesting a negligible contribution of familiarity to FCNC. Despite significant zero-order correlations between FCNC and each of these three variables, none are significant unique predictors of FCNC in the full model. This indicates that the correlations of these variables with FCNC are driven primarily by variance shared by all three predictors, which is consistent with a common role of recollection in all three tasks.

When the same analysis was performed for FCC, the proportion of variance explained by recall measures in Stage 1 was not significant [$R^2 = .055$, $F(2, 65) = 1.909$, $p = .156$]. This suggests that, unlike FCNC, recollection played a negligible role in FCC performance. However, the addition of the recognition measure at Stage 2 significantly increased the variance explained by the model [$\Delta R^2 = .120$, $F(1, 64) = 9.303$, $p = .003$]. This suggests that FCC performance was driven largely by familiarity. Importantly, although similar total proportions of variance in FCC and FCNC were explained by Stage 2 (R^2 values of .192 and .175, respectively), with the same standardized recall and recognition tests as predictors, it was attributable to different predictors in each case. This is the pattern of results expected if there is the selective engagement of familiarity in FCC and recollection in FCNC performance.

Table 2. *Correlation matrix of tests included in hierarchical regression analysis*

				Doors & People		*WMS–IIIA*			
		FCC	*FCNC*	*Recall*	*Recognition*	*Immediate*	*Delayed*	*Total*	*Predicted FSIQ*
Measure		r	r	r	r	r	r	r	r
FCNC		.339 **							
Doors & People	Recall	.199	.357 **						
	Recognition	.411 **	.366 **	.507 ***					
WMS–IIIA	Immediate	.163	.269 *	.418 ***	.335 **				
	Delayed	.251 *	.336 **	.502 ***	.340 **	.864 ***			
	Total	.205	.312 *	.476 ***	.345 **	.963 ***	.967 ***		
Predicted FSIQ		.168	.270 *	.286 *	.188	.258 *	.252 *	.271 *	
Age		−.248 *	−.137	−.028	−.117	.030	.015	.025	.333 **

Note: All values are Pearson's correlation coefficients to three decimal places. FCC = forced-choice corresponding; FCNC = forced-choice noncorresponding; WMS–IIIA= Wechsler Memory Scale–Third Edition Abbreviated; FSIQ = full-scale intelligence quotient.
*p < 05. **p < .01. ***p < .001.

Table 3. *Multiple regression results*

Model	*Step*	*Variable*	*B*	SE *B*	*β*	*Adjusted R*^{2}
FCNC model						
	Step 1					.128
		Constant	0.088	0.206		
		WMS–IIIA total	0.003	0.002	0.184	
		DP recall	0.017	0.008	0.269*	
	Step 2					.154
		Constant	−0.001	0.206		
		WMS–IIIA total	0.003	0.002	0.153	
		DP recall	0.011	0.009	0.168	
		DP recognition	0.017	0.010	0.228	
FCC model						
	Step 1					.026
		Constant	0.399	0.187		
		WMS–IIIA total	0.002	0.002	0.143	
		DP recall	0.007	0.008	0.131	
	Step 2					.137
		Constant	0.263	0.181		
		WMS–IIIA total	0.001	0.002	0.088	
		DP recall	−0.003	0.008	−0.048	
		DP recognition	0.026	0.008	0.405**	

Note: FCC = forced-choice corresponding; FCNC = forced-choice noncorresponding; WMS–IIIA= Wechsler Memory Scale–Third Edition Abbreviated; DP = Doors & People Test.
*$p < .05$. **$p < .01$.

Simulation results

Latent factor simulations were conducted to determine how often we should expect to observe the same arrangement of hierarchical regression outcomes if the data are known to be generated from a single memory process compared to two memory processes. Factor models of recall and recognition representing single-process or dual-process assumptions (e.g., Quamme et al., 2004; Unsworth & Brewer, 2009) were used for the simulation. We created synthetic datasets from single- and dual-process factor models and subjected them to the same hierarchical regression analyses as those reported above in order to see how often we should expect the same regression outcomes if either model was actually true.

A factor model assumes that measured variables correlate with one another across individuals to the degree they load on common underlying factors. A single memory process can be represented as a single-factor model, in which recall, recognition, and the forced-choice measures load on one common factor, but to varying degrees (see Figure 3a). The dual-process model has two independent memory factors, representing recollection and familiarity. The recall measures and FCNC load on only the recollection factor, whereas DP recognition and FCC load on both the recollection and familiarity factors (see Figure 3b). This means that that whereas both FCC and FCNC measures load on recollection (albeit unequally), FCC loads on familiarity, and FCNC does not. Although WMS–III total and DP recall did not predict FCC significantly in the regression analysis, the theoretical account gives no a priori reason to expect recollection to be strictly absent from FCC (unlike the case of familiarity's absence from FCNC). We therefore allowed FCC to load on recollection in the factor model. If anything, this should work against reproducing the observed regression results, because it should increase the correlation between FCC and recall measures, thus increasing the chances of a significant R^2 in Stage 1 for FCC.

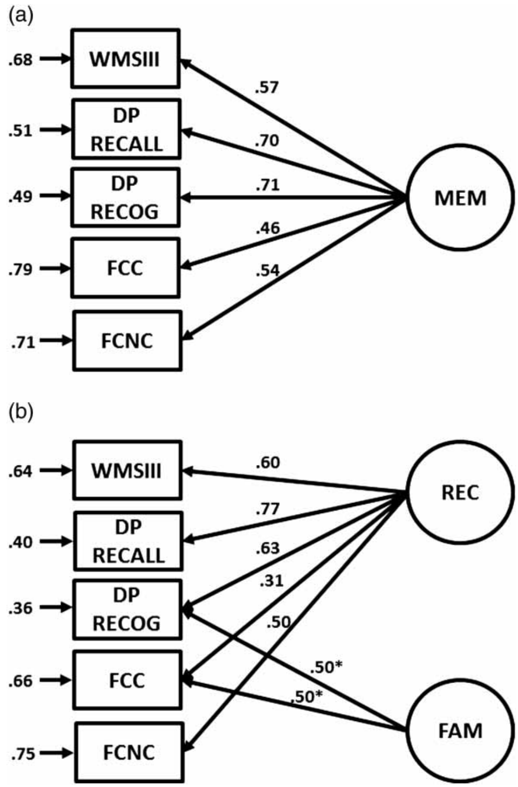

Figure 3. *Factor structure and standardized parameter values for the single-process model (Panel A) and dual-process model (Panel B) used in the simulations. Factor variances were set to 1.0. Fixed parameter values are denoted by *; all other parameters were freely estimated to find the best fitting value. WMSIII = Wechsler Memory Scale–Third Edition; DP = Doors and People Test; recog = recognition; FCC = forced-choice corresponding; FCNC = forced-choice noncorresponding; MEM = memory factor; REC = recollection factor; FAM = familiarity factor.*

The simulation procedure followed five basic steps. First, single- and dual-process models were fitted to the observed data to obtain plausible parameter values. Secondly, using these model parameters, data from 10,000 samples of 68 synthetic subjects each were created. Thirdly, the recall, recognition, and forced-choice measures of each sample were analysed using the same hierarchical regression procedures as those in the experiment. Fourthly, each sample was inspected to determine whether the regression analyses reproduced each of the four regression outcomes reported in the observed data: (a) nonsignificant R^2 at Stage 1 for FCC; (b) significant R^2 at Stage 2 for FCC; (c) significant R^2 at Stage 1 for FCNC; and (d) nonsignificant R^2 at Stage 2 for FCNC. Finally, the number of samples was tallied in each simulation for which all four observed regression outcomes were reproduced.

Model parameters. The parameters for the factor models consisted of factor loadings (one parameter value per measure per factor), which account for shared variance among measures, and error variances (one per measure), which account for unshared variance unique to each measure. To obtain plausible parameter values for the simulation, the models were fitted by maximum likelihood estimation to the covariance matrix using LISREL 8.8. For the single–process model (Figure 3a), the five loadings and five error variances were freely estimated. The fit was not rejectable, indicating that this model provided a plausible account of the data [$\chi^2(5) = 7.12$, $p = .21$; root mean square error of approximation, RMSEA = .079; non-normed fit index, NNFI = 0.95; comparative fit index, CFI = .97]. For the dual-process model (Figure 3b), there was no unique best fitting solution when all parameters (five recollection loadings, five error variances, and two familiarity loadings) were estimated. The familiarity factor was estimated from only two measures, and three measures per factor are typically needed unless restrictions are placed on loadings and/or residual values. Familiarity loadings for DP recognition and FCC were therefore arbitrarily fixed at standardized values of .5, while the other parameters were estimated as in the single-process model. This means that not all the dual-process model parameters are optimized for best fit. However, the fit was numerically better than that of the single-process model, indicating that the dual-process parameters are at least as plausible as those of the single-process model [$\chi^2(5) = 4.61$, $p = .47$, RMSEA = .00, NNFI = 1.01, CFI = 1.0]. Note that the single- and dual-process models technically have the same free parameters and can be understood as two versions of the same model. For the single-process model, familiarity loadings are all simply fixed at 0, whereas the dual-process model has two of the familiarity loadings fixed at a standardized value of .5 and the other three fixed at 0. Otherwise, all the same loadings and error variances are estimated.

Simulation procedures. Data for synthetic subjects were created by randomly sampling values from normal distributions ($\mu = 0$, $\sigma = 1$) to serve as factor scores (one score per factor) and error scores (one per test) for each subject. These values were used along with the model parameters to calculate synthetic observed scores for the five memory tests. Six values per subject were generated for the single-process model: one for the memory factor (MEM) and five residual error scores (RES_i) for the five memory tests. For the dual-process model, seven values were generated: one each for the recollection (REC) and familiarity (FAM) factors, and the five residual scores for the five memory tests. Observed scores for the memory tests were then calculated for each simulated subject from the sampled factor and residual scores using the parameters of the model. For the single-process model, the observed score for a subject on a given test i ($TESTSCORE_i$) is calculated as:

$$TESTSCORE_i = \lambda_i \times MEM + e_i \times RES_i$$

where λ_i is the loading of test i on the memory factor (MEM), and e_i is the square root of the error variance for test i.

For the dual-process model, the observed test scores are calculated as:

$$TESTSCORE_i = \lambda_{ir} \times REC + \lambda_{if} \times FAM + e_i \times RES_i$$

Where λ_{ir} is the recollection (REC) loading, λ_{if} is the familiarity (FAM) loading, and e_i is again the square root of the error variance for test i.

In this way, 10,000 samples of 68 synthetic subjects were generated from each of the two models. Each subject record in each sample contained simulated versions of the five memory measures used in the experiment. See Supplemental Material for more details.

Analysis and results of simulated data. Hierarchical regression analyses were run on each of the samples to determine how often, under the assumptions of each model, we would expect to encounter the four statistical outcomes observed in the real data described earlier. Out of 10,000 samples, the single-process model reproduced this outcome 335 times (or with .0335 probability), whereas the dual-process model reproduced it 3,505 times (or with .3505 probability). In other words, the joint occurrence of the four observed regression outcomes in our experiment is more probable by approximately a factor of 10 under the dual- than under the single-process model, given our sample size.

A number of additional simulations were conducted to further understand these results. The dual-process model only generated the observed results about one third of the time, despite doing so 10 times as frequently as the single-process model. There are several possible reasons for this. First, the familiarity loadings of the dual-process model are not optimized for best fit; an optimized solution may have resulted in a greater success rate. Secondly, the FCC recollection loading works against reproducing the observed result by increasing the correlations between recall and FCC, thus increasing the chances of a significant R^2 in Stage 1 for FCC. Running the simulation with the FCC recollection loading removed (i.e., set to 0), increased reproduction rate of the observed regression results to 4947/10,000, or almost 50%.

A third potential reason is that reproducing the observed results requires joint reproduction of four separate outcomes, so the sample size may be underpowered to detect all components of this pattern at once. It is reasonable to ask whether both models would produce the observed results more often if sample size were greater. However, this is not the case; when the original simulations are re-run with sample sizes of 200 and 500, the reproduction rate for the single-process model dropped to 152/10,000 with 200 subjects and to

31/10,000 with 500 subjects. The reproduction rates for the dual-process model increased to 6122/10,000 with 200 subjects (approximately 60%) and to 8052/10,000 with 500 subjects (approximately 80%). Increasing sample size helps to differentiate the two models further, with our outcome becoming much more common under the dual-process model (despite nonoptimal familiarity parameters and a nonzero FCC recollection loading) and much less likely under the single-process model.

We also examined other two-factor models in which the factors did not correspond to recollection and familiarity. Of particular relevance are models with a single memory process underlying recall and recognition (e.g., a single memory factor on which all tests load), but with a second factor contributing to two or more tests representing something other than familiarity. One such possibility is a *dual-process recall* model, in which a second factor contributes only to recall, representing additional memory search or response generation demands of recall performance not shared by recognition (Quamme et al., 2004). Another possibility is a *stimulus category factor* model, in which the second factor represents shared variance among FCC and FCNC associated with the use of the same category of stimuli (object pictures) in these two tests. To simulate both of these models, the same procedure was used as that for the dual-process model, where the loadings of two measures on a second factor were fixed to have a standardized value of .5 (the two recall measures for dual-process recall, FCC, and FCNC measures for the stimulus category factor). The other measures did not load on the second factor, and the remaining parameters were estimated by fitting the model. Out of 10,000 samples, the expected outcome occurred for the dual-process recall model only 380 times and for the stimulus category factor model only 315 times. Neither of these models therefore achieved an appreciatively greater success rate than the single-process model. This shows that it is the additional shared variance between recognition and FCC specifically, as predicted by the CLS's dual-process account, which makes the observed regression results likely to occur.

Discussion

In this study, a group of older adults completed standardized neuropsychological tests to assess memory and IQ, as well an experimental memory test requiring participants to distinguish target pictures from very similar foils in two different test formats (FCC and FCNC). The standardized memory tests gave independent estimates of recall and recognition memory performance. These allowed us to test for patterns of shared and unshared variance in performance across individuals, consistent with differential contributions of recollection and familiarity, as predicted by the CLS model (Norman & O'Reilly, 2003). FCNC, but not FCC, was significantly predicted by standardized measures of recall. This suggests that a recollection-related variance partition was present in FCNC, but absent/minimal in FCC. In contrast, once variance shared by recall was accounted for, the remaining variance in FCC but not FCNC was predicted by recognition. This suggests that a familiarity-related variance partition was present in FCC, but absent/minimal in FCNC.

To confirm this interpretation, a simulation analysis was conducted. Here, we generated simulated data from a dual-process model and from alternative single- and two-process models. We then performed the same regression analyses on the synthetic data. The simulations reveal three important things about our observed regression results. First, the results are expected about 10 times more frequently under the dual-process account than under the best fitting single-process account. Secondly, the results are not simply artefacts of low power: The observed outcomes become *more* likely for the dual-process model and *less* likely for the single-process model as sample size increases. Finally, it is not the case that just any two-factor account of the data will do: The regression outcomes are 10 times less frequent under other plausible two-factor models than in the dual-process account we advocate here. Across all of the simulations carried out, our observed results were only reproduced often in models including both familiarity and recollection factors. When using all the other possible models, our observed results were reproduced at rates of less than 5%. We therefore conclude that the observed regression results are diagnostic of a relatively specific pattern of individual differences that is consistent with CLS expectations and inconsistent with expectations of a single memory process.

The present results corroborate the predictions of the CLS model (Norman & O'Reilly, 2003) with novel evidence from an individual differences approach for a differential contribution of recollection and familiarity to FCC and FCNC performance. As explained in the introduction, good FCC performance can rely on familiarity but good FCNC performance requires recollection. When applied to individual differences, these predictions translate to the expectation that a shared contribution of recollection, but not familiarity, should in part drive the correlations between FCNC and other kinds of memory performance. In contrast, a shared contribution of familiarity should in part drive such correlations for FCC. The present study explicitly confirms this expectation, under the assumption that standardized tests of recall are viable proxies of a common recollection process and that DP recognition at least in part relies on familiarity.

Note that these assumptions do not *equate* tasks with processes, such as equating recall with recollection. It is only assumed that shared variance among two tasks is the result of one or more common processes. The assumptions of the present study are consistent with the widespread view that recall and recollection-based recognition rely on common cognitive and neural processes (for a review of theories of recollection see Moulin, Souchay, & Morris, 2013). Recall and recognition may also share additional processes other than those strictly attributable to recollection (McCabe & Soderstrom, 2011; Mickes, Seale-Carlisle, & Wixted, 2013). However, unless these processes behave the same way as we predicted recollection to behave (and not familiarity), they do not explain why divergent variance partitions were found in FCC and FCNC.

Although we did not have strong predictions about the role of recollection in FCC, the data are actually most consistent with a negligible contribution from recollection. The simulation results show that the observed regression outcomes are actually more expected if FCC depends only on familiarity than if it contains a modest amount of recollection. We hasten to add that the present results should not be taken as evidence of strict process purity for either FCC or FCNC. However, the results suggest a strongly disproportionate (and roughly symmetrical) weighting of FCC toward familiarity and FCNC toward recollection.

The results also converge with prior studies of FCC and FCNC performance in both young healthy individuals and amnesic patients. Patients with hippocampal damage (Holdstock et al., 2002) and aMCI (Westerberg et al., 2013; Westerberg et al., 2006), who show recall deficits and impaired YN recognition performance with high target–foil similarity, show sparing of FCC performance. On the other hand, patients with impairments in both recollection and familiarity perform poorly on FCC, FCNC and YN formats (Jeneson, Kirwan, Hopkins, Wixted, & Squire, 2010). Young healthy participants instructed to use only familiarity to make recognition judgements show a reduction in FCNC performance but not FCC performance compared to standard recognition instructions (Migo et al., 2009). These observations are all consistent with the current finding in suggesting that FCC performance is largely familiarity based, whereas FCNC performance is largely recollection based.

Although in this study recollection and familiarity were not measured directly, our assumptions about how these processes contribute to standard recall and recognition measures have been tested directly in prior studies of individual differences in memory performance. Models in which both recall and recognition load on a common factor (recollection), while only recognition loads on a second factor (familiarity), have consistently been shown to provide better fits than models assuming only one factor or different two-factor structures (Quamme et al., 2004; Unsworth & Brewer, 2009; Vann et al., 2009; Yonelinas et al., 2007).

The current study is consistent with, and complementary to, prior latent-factor studies of recollection and familiarity. The present experiment did not provide enough data to fit the models in a way that would make them statistically comparable. However, unlike prior latent-factor studies, the present conclusions are based on simulation performance, rather than model fitting. The current study adds to the previous factor literature by showing that the dual-process factor structure is associated with a characteristic pattern of shared variance amongst memory measures that is detectable with simpler regression methods. There may be alternative explanations of the present findings, but these must explain why recall performance should successfully predict FCNC but not FCC performance, whereas additional variance in recognition, above and beyond that shared by recall, should selectively predict FCC, but not FCNC performance. The dual-process recollection/familiarity account given by the CLS model, on the other hand, explicitly makes this prediction.

From a practical assessment perspective, previous work has suggested that FCC versus FCNC comparisons are more appropriate than those comparing FCC and YN (Migo et al., 2009). CLS predictions for FCNC and YN recognition under conditions of high target–foil similarity are essentially the same. The forced-choice format, however, is less affected by response bias than YN (Macmillan & Creelman, 2004), and direct comparisons of YN and FCC confound test length with test format. FCC and FCNC tests have identical numbers of trials, removing any potential confound over differing test length. The combined FCC/FCNC test used here could be practically useful in neuropsychological test design, as a single recognition memory test that differentially requires recollection for success on certain trials. It has simple instructions and can assess recollection-dependent performance without the need for free recall trials, which can be stressful for participants with very poor memories as being repeatedly exposed to failure can be demotivating (Morris, 2004). Other ways to assess recollection and familiarity performance such as the remember/know procedure (Migo, Mayes, & Montaldi, 2012; Tulving, 1985) or the process dissociation procedure (Jacoby, 1991; Yonelinas & Jacoby, 2012) require complicated instructions and/or experimental designs. This makes a single FCC/FCNC test especially appropriate for clinical or older populations. Older adults did not report having any problems completing the task in this study.

The present data highlight the importance of task design in assessing the role of recollection and familiarity in recognition. It is not appropriate to assume that all recognition tests can be supported by familiarity, since we have two recognition tests here showing opposite patterns. The assumption that some other test formats, such as associative recognition or source memory tasks, necessarily require recollection has equally been challenged (e.g., Migo et al., 2012). The specific test materials and method of presentation at test are important factors to consider, along with other established influences, such as list length or divided attention.

Our use of older adults helped to ensure that standardized test performance and recollection levels were not at ceiling. Recollection declines with age (e.g., Bastin & Van der Linden, 2003; Bugaiska et al., 2007; Howard, Bessette-Symons, Zhang, & Hoyer, 2006; Light, Prull, La Voie, & Healy, 2000; McCabe, Roediger, McDaniel, & Balota, 2009; Prull, Dawes, Martin, Rosenberg, & Light, 2006), whereas familiarity appears relatively stable (e.g., Friedman, 2013; Howard et al., 2006; Light et al., 2000; McCabe et al., 2009; Parkin & Walter, 1992). By testing a sample of older adults, we ensured there would be sufficient variance in standardized test performance for correlation and regression analyses. We also reasoned that older adults would be generally less likely than student subjects to use recollection and instead use familiarity more often when possible. A sample of students might have reduced our ability to examine patterns of shared variance among the variance memory measures. Younger subjects' greater recollection ability may also have reduced our ability to test for familiarity-related variance.

One perhaps surprising result from this study is the lack of significant correlations between the FCNC test and age. The decline of recall/recollection with age is well established but this pattern was not seen for FCNC, although there was a significant negative correlation between age and FCC. It is important to note that most studies showing an age related decline in recollection compare groups of younger and older adults, rather than looking for within-group correlations between memory performance and age. Where that analysis has been carried out, effect sizes remain small (e.g., a mean correlation of −.13 between age and recall in 364 older adults aged 65 to 80 years in Zimprich & Kurtz, 2013). Studies have shown that there are unique age contributions to memory, but these remain small after accounting for a common contributor to cognition or a measure of executive function (e.g., Krueger & Salthouse, 2011; Rhodes & Kelley, 2005; Salthouse, 2009). Although there are clearly overall differences between older and younger adults on measures of memory, the evidence suggests that on an individual-by-individual basis, age alone is not a good predictor of performance. Our significant positive correlation between age and IQ indicates that our sample has some very high IQ scoring older adults at the top of our age range. This relationship makes any further interpretations of the relationship between FCC/FCNC and age/IQ in the present data problematic. Given that an investigation of ageing effects was not our primary aim, and that we did not collect data on other cognition functions or attempt to match IQ across the ages sampled, we are unable to explore this further in our data.

Summary

The results provide convergent evidence that forced-choice recognition performance with high target–foil similarity depends on distinct memory processes when tested in different formats. We have shown that in this group of older adults, FCNC performance was dependent on recollection but FCC was dependent on familiarity. We have demonstrated this using correlations, hierarchical regression, and a factor simulation procedure. This finding is consistent with dual-process interpretations of recognition memory and confirms the specific predictions of the CLS neurocomputational model.

Supplemental material

Supplemental content is available via the "Supplemental" tab on the article's online page (http://dx.doi.10.1080/17470218.2014.910240.2014.PQJE910240).

Original manuscript received 5 April 2013
Accepted revision received 18 March 2014
First published online 6 May 2014

REFERENCES

Anderson, N. D., Ebert, P. L., Jennings, J. M., Grady, C. L., Cabeza, R., & Graham, S. J. (2008). Recollection- and familiarity-based memory in healthy aging and amnestic mild cognitive impairment. *Neuropsychology*, *22*(2), 177–187. doi:10.1037/0894–4105.22.2.177

Baddeley, A. D., Emslie, H., & Nimmo-Smith, I. (1994). *Doors and people: A test of visual and verbal recall and recognition*. Bury St. Edmunds: Thames Valley Test Co.

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology*, *17*(1), 14–24. doi:10.1037/0894–4105.17.1.14

Belleville, S., Ménard, M. C., & Lepage, É. (2011). Impact of novelty and type of material on recognition in healthy older adults and persons with mild cognitive impairment. *Neuropsychologia*, *49*(10), 2856–2865. doi:10.1016/j.neuropsychologia.2011.06.011

Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus?. *Nature Reviews Neuroscience*, *2*(1), 51–61. doi:10.1038/35049064

Bugaiska, A., Clarys, D., Jarry, C., Taconnat, L., Tapia, G., Vanneste, S., & Isingrini, M. (2007). The effect of aging in recollective experience: the processing speed and executive functioning hypothesis. *Consciousness and Cognition*, *16*(4), 797–808. doi:10.1016/j.concog.2006.11.007

Deason, R. G., Hussey, E. P., Budson, A. E., & Ally, B. A. (2012). Gist-based conceptual processing of pictures remains intact in patients with amnestic mild cognitive impairment. *Neuropsychology*, *26*(2), 202–208. doi:10.1037/a0026958

Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., & Lipton, P. (2012). Towards a functional organization of episodic memory in the medial temporal lobe. *Neuroscience and Biobehavioral Reviews*, *36*(7), 1597–1608.

Eichenbaum, H., Yonelinas, A. R., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*(1), 123–152. doi:10.1146/annurev.neuro.30.051606.094328

Embree, L. M., Budson, A. E., & Ally, B. A. (2012). Memorial familiarity remains intact for pictures but not for words in patients with amnestic mild cognitive impairment. *Neuropsychologia*, *50*(9), 2333–2340. doi:10.1016/j.neuropsychologia.2012.06.001

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Friedman, D. (2013). The cognitive aging of episodic memory: A view based on the event-related brain potential. *Frontiers in Behavioral Neuroscience* (AUG). doi:10.3389/fnbeh.2013.00111

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods Instruments & Computers*, *26*(4), 381–386.

Hintzman, D. L., Curran, T., & Oppy, B. (1992). Effects of similarity and repetition on memory: registration without learning?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 667–680.

Holdstock, J. S., Mayes, A. R., Roberts, N., Cezayirli, E., Isaac, C. L., O'Reilly, R. C., & Norman, K. A. (2002). Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans?. *Hippocampus*, *12*(3), 341–351. doi:10.1002/hipo.10011

Howard, M. W., Bessette-Symons, B., Zhang, Y., & Hoyer, W. J. (2006). Aging selectively impairs recollection in recognition memory for pictures: Evidence from modeling and receiver operating characteristic curves. *Psychology and aging*, *21*(1), 96–106. doi:10.1037/0882–7974.21.1.96

Hudon, C., Belleville, S., & Gauthier, S. (2009). The assessment of recognition memory using the Remember/Know procedure in amnestic mild cognitive impairment and probable Alzheimer's disease. *Brain and Cognition*, *70*(1), 171–179. doi:10.1016/j.bandc.2009.01.009

Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. doi:10.1016/0749–596X(91)90025-F

Jeneson, A., Kirwan, C. B., Hopkins, R. O., Wixted, J. T., & Squire, L. R. (2010). Recognition memory and the hippocampus: A test of the hippocampal contribution to recollection and familiarity. *Learning & Memory*, *17*(1), 63–70. doi:10.1101/lm.1546110

Krueger, L. E., & Salthouse, T. A. (2011). Influence of cognitive abilities and age on word recall performance across trials and list segments. *American Journal of Psychology*, *124*(3), 291–300. doi:10.1016/j.intell.2004.07.005

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

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Mandler, G. (1980). Recognizing: the judgement of previous occurrance. *Psychological Review*, *87*(3), 252–271. doi:10.1037/0033–295X.87.3.252

McCabe, D. P., Roediger, H. L., McDaniel, M. A., & Balota, D. A. (2009). Aging reduces veridical remembering but increases false remembering: Neuropsychological test correlates of remember–know judgments. *Neuropsychologia*, *47*(11), 2164–2173. doi:http://dx.doi.org/10.1016/j.neuropsycholo gia.2008.11.025

McCabe, D. P., & Soderstrom, N. C. (2011). Recollection-based prospective metamemory judgments are more accurate than those based on confidence: Judgments of remembering and knowing (JORKs). *Journal of Experimental Psychology: General*, *140*(4), 605–621. doi:10.1037/a0024014

Mickes, L., Seale-Carlisle, T. M., & Wixted, J. T. (2013). Rethinking familiarity: Remember/Know judgments in free recall. *Journal of Memory and Language*, *68*(4), 333–349. doi:10.1016/j.jml.2013.01.001

Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Consciousness and Cognition*, *21*(3), 1435–1455. doi:10.1016/j.concog.2012.04.014

Migo, E. M., Montaldi, D., & Mayes, A. R. (2013). A visual object stimulus database with standardized similarity information. *Behavior Research Methods*, *45*(2), 344–354. doi:10.3758/s13428–012–0255–4

Migo, E. M., Montaldi, D., Norman, K. A., Quamme, J. R., & Mayes, A. R. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *Quarterly Journal of Experimental Psychology*, *62*(6), 1198–1215. doi:10.1080/17470210802391599

Migo, E. M., Weiss, D., Norman, K., Mayes, A., & Montaldi, D. (2008, April). *A new collection of fifty picture sets with standardised similarity information*. Paper presented at the Experimental Psychology Society, Cambridge.

Montaldi, D., & Mayes, A. R. (2010). The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus*, *20*(11), 1291–1314. doi:10.1002/hipo.20853

Morris, R. G. (2004). Neuropsychology of older adults. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians* (pp. 301–318). Chichester, West Sussex, UK: Wiley.

Moulin, C. J. A., Souchay, C., & Morris, R. G. (2013). The cognitive neuropsychology of recollection. *Cortex*, *49*(6), 1445–1451. doi:http://dx.doi.org/10.1016/j.cortex.2013.04.006

Norman, K. A. (2010). How hippocampus and cortex contribute to recognition memory: Revisiting the complementary learning systems model. *Hippocampus*, *20*(11), 1217–1227. doi:10.1002/hipo.20855

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. doi:10.1037/0033–295X.110.4.611

O'Connor, M. K., & Ally, B. A. (2010). Using stimulus form change to understand memorial familiarity for pictures and words in patients with mild cognitive impairment and Alzheimer's disease. *Neuropsychologia*, *48*(7), 2068–2074. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2010.03.027

Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and aging*, *7*(2), 290–298. doi:10.1037//0882–7974.7.2.290

Prull, M. W., Dawes, L. L. C., Martin, A. M., Rosenberg, H. F., & Light, L. L. (2006). Recollection and familiarity in recognition memory: Adult age differences and neuropsychological test correlates. *Psychology and aging*, *21*(1), 107–118. doi:10.1037/0882–7974.21.1.107

Quamme, J., Yonelinas, A. P., Widaman, K. F., Kroll, N. E., & Sauve, M. J. (2004). Recall and recognition in mild hypoxia: Using covariance structural modeling to test competing theories of explicit memory. *Neuropsychologia*, *42*(5), 672–691. doi:10.1016/j.neuropsychologia.2003.09.008

Ranganath, C. (2010). A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus*, *20*(11), 1263–1290. doi:10.1002/hipo.20852

Rhodes, M. G., & Kelley, C. M. (2005). Executive processes, memory accuracy, and memory monitoring: An aging and individual difference analysis. *Journal of Memory and Language*, *52*(4), 578–594. doi:http://dx.doi.org/10.1016/j.jml.2005.01.014

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: evidence from ROC curves. *Journal of Memory and Language*, *43*(1), 67–88. doi:10.1006/jmla.1999.2701

Salthouse, T. A. (2009). Decomposing age correlations on neuropsychological and cognitive variables. *Journal of the International Neuropsychological Society*, *15*(5), 650–661.

Schefter, M., Werheid, K., Almkvist, O., Lönnqvist-Akenine, U., Kathmann, N., & Winblad, B. (2013). Recognition memory for emotional faces in amnestic mild cognitive impairment: An event-related potential study. *Aging, Neuropsychology, and Cognition*, *20*(1), 49–79. doi:10.1080/13825585.2012.665021

Serra, L., Bozzali, M., Cercignani, M., Perri, R., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2010). Recollection and familiarity in amnesic mild cognitive impairment. *Neuropsychology*, *24*(3), 316–326. doi:10.1037/a0017654

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 479–496.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1–12. doi:10.1037/h0080017

Unsworth, N., & Brewer, G. A. (2009). Examining the relationships among item recognition, source recognition, and recall from an individual differences perspective. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(6), 1578–1585. doi:10.1037/a0017255

Vann, S. D., Tsivilis, D., Denby, C. E., Quamme, J. R., Yonelinas, A. P., Aggleton, J. P., … Mayes, A. R. (2009). Impaired recollection but spared familiarity in patients with extended hippocampal system damage revealed by 3 convergent methods.

Proceedings of the National Academy of Sciences, *106*(13), 5442–5447. doi:10.1073/pnas.0812097106

Wais, P. E. (2013). The limited usefulness of models based on recollection and familiarity. *Journal of Neurophysiology*, *109*(7), 1687–1689. doi:10.1152/jn.00548.2012

Wechsler, D. (1998). *Wechsler Memory Scale® - 3rd UK Edition*. London: The Psychological Corporation.

Wechsler, D. (2001). *Wechsler test of adult reading*. Oxford: Harcourt Assessments.

Westerberg, C. E., Mayes, A., Florczak, S. M., Chen, Y., Creery, J., Parrish, T., … Paller, K. A. (2013). Distinct medial temporal contributions to different forms of recognition in amnestic mild cognitive impairment and Alzheimer's disease. *Neuropsychologia*, *51*(12), 2450–2461. http://dx.doi.org/10.1016/j.neuropsychologia.2013.06.025

Westerberg, C. E., Paller, K. A., Weintraub, S., Mesulam, M. M., Holdstock, J. S., Mayes, A. R., & Reber, P. J. (2006). When memory does not fail: Familiarity-based recognition in mild cognitive impairment and Alzheimer's disease. *Neuropsychology*, *20*(2), 193–205. doi:10.1037/0894–4105.20.2.193

Yonelinas, A. P., & Jacoby, L. L. (2012). The process-dissociation approach two decades later: Convergence, boundary conditions, and new directions. *Memory and Cognition*, *40*(5), 663–680. doi:10.1016/j.neuropsychologia.2005.01.019

Yonelinas, A. P., Widaman, K., Mungas, D., Reed, B., Weiner, M. W., & Chui, H. C. (2007). Memory in the aging brain: Doubly dissociating the contribution of the hippocampus and entorhinal cortex. *Hippocampus*, *17*(11), 1134–1140. doi:10.1002/hipo.20341

Zimprich, D., & Kurtz, T. (2013). Individual differences and predictors of forgetting in old age: The role of processing speed and working memory. *Aging, Neuropsychology, and Cognition*, *20*(2), 195–219.