Aging selectively impairs recollection in recognition memory for pictures: Evidence from modeling and ROC curves

Marc W. Howard, Brandy Bessette-Symons, Yaofei Zhang, William J. Hoyer
Department of Psychology
Syracuse University

In press, *Psychology and Aging*

Younger and older adults were tested on recognition memory for pictures. The Yonelinas high threshold (YHT) model, a formal implementation of two-process theory, fit the response distribution data of both younger and older adults significantly better than a normal unequal variance signal detection model. Consistent with this finding, non-linear zROC curves were obtained for both groups. Estimates of recollection from the YHT model were significantly higher for younger than older adults. This deficit was not a consequence of a general decline in memory; older adults showed comparable overall accuracy and in fact a non-significant increase in their familiarity scores. Implications of these results for theories of recognition memory and the mnemonic deficit associated with aging are discussed.

Recognition memory tests the abilities that enable us to identify an old friend in a sea of faces, or to know that we have in fact heard that joke before. In laboratory studies of recognition memory, participants are presented with a list of to-be remembered items. During a test phase the participant is presented with a series of probe items, some of which were present in the list and some of which are new items. The participant's task is to determine which are which.

Numerous theorists have argued that recognition memory is not a unitary ability, but is rather composed of two processes (Atkinson & Juola, 1974; Mandler, 1980; Tulving, 1983; Yonelinas, 1997). In the modal version of this idea, one component, referred to as recollection, describes our ability to recover vivid detailed information about the study episode. Recollection provides strong evidence of prior occurrence that enables us to endorse recollected old probe items with certainty. In addition, two-process theory postulates that familiarity, a memory process that corresponds to enhanced perceptual or semantic fluency for recently-experienced items also supports recognition performance even in the absence of vivid recollection. Researchers have developed a variety of experimental techniques designed to measure recollection and familiarity, including the process dissociation procedure (Jacoby, 1991), the remember/know procedure (Tulving, 1985), tests of source memory (Dywan & Jacoby, 1990) and the analysis of receiver operating characteristic (ROC) curves (Yonelinas, 1994, 1997). Numerous experimental dissociations between recollection and familiarity have been observed. For instance, recollective detail becomes available later than familiarity in retrieval (Hintzman & Curran, 1994) and depends on the integrity of the hippocampus (Yonelinas et al., 2002; Fortin, Wright, & Eichenbaum, 2004).

Aging and two-process theory.

Changes in memory function with normal aging have been studied from the perspective of two-process theory using a variety of techniques. The consensus from these studies is that aging affects recollection. Insofar as recollection is similar to recall and age deficits in recall have been extensively observed (e.g., Naveh-Benjamin, 2000; Kahana, Howard, Zaromb, & Wingfield, 2002; Onyper, Hoyer, & Cerella, in press), this is not surprising. There is less consensus as to whether the deficit with aging is specific to recollection or if aging also results in a decrease in familiarity as well (Hoyer & Verhaeghen, In press; Prull, Crandell-Dawes, Martin, Rosenberg, & Light, In press).

In associative recognition participants study pairs of items (e.g., A-B, C-D) and must distinguish old pairs from repaired lures (e.g., A-D). Because each of the components of a re-paired lure should be familiar, association recognition is believed to rely preferentially on recollection. Some have argued that aging impairs associative recognition more than item recognition, which presumably depends on both recollection and familiarity (Light, Patterson, Chung, & Healy, 2004; Naveh-Benjamin, 2000; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003; Naveh-Benjamin, Guez, Kilb, & Reedy, 2004). In particular, Light et al. (2004) showed that associative recognition performance for older adults was comparable to that observed for younger adults given only a short amount of time to respond. Given the finding that recollective information is retrieved more slowly than information about familiarity (Hintzman & Curran, 1994, see also

The authors wish to acknowledge MH069938 to MWH and AG11451 to WJH. The authors thank Marty Sliwinski for his invaluable advice on some of the analyses and Greg Mangan for his assistance with data collection. Send correspondence to Marc Howard, Syracuse University, 430 Huntington Hall, Syracuse, NY 13244-2340, marc AT memory DOT syr DOT edu.

Rotello & Heit, 1999) this is especially suggestive of a strong recollective deficit in older adults.

Parks, Toth, and Smith (2004) estimated recollection and familiarity for younger and older adults using the remember-know procedure for memory of words presented in various voices and fonts. On the basis of their findings, they argued that normal aging was associated with reductions in both recollection and familiarity. In contrast, other remember-know studies (e.g., Bastin & Van Der Linden, 2003) have argued that the mnemonic deficit associated with aging is specific to recollection, with familiarity slightly higher for older adults. The potential lack of independence of recollection and familiarity make it particularly difficult to estimate recollection and familiarity from remember-know judgments, even if one takes the remember-know procedure at face value in directly measuring categorically different responses (an assumption that may not be viable, see Wixted & Stretch, 2004). Another criticism of the remember-know procedure is that remember and know judgements are not process-pure insofar as sufficiently strong familiarity for an item may lead to a remember judgement.

Other methods for assessing recollection and familiarity have been used to assess the mnemonic deficit in normal aging. Reasoning that recall depends on recollection but recognition relies on both recollection and familiarity, Quamme, Yonelinas, Widaman, Kroll, and Sauve (2004) estimated recollection and familiarity from structural equation modeling of recall and recognition scores for adults that varied in age and hypoxic damage. The results of the structural equation modeling argued that age affected recollection but not familiarity. Similarly, Jennings and Jacoby (1997) suggested on the basis of process dissociation procedure data that older adults were impaired at recollection while relatively unaffected in their automatic memory processes, i.e. familiarity.

Modeling and ROC curves as a constraint on theory

Although there are many ways to assess recollection and familiarity (for a review see Yonelinas, 2002), in this paper we will focus on fitting models of item recognition to ROC curves. Measures of recognition memory describe discriminability as a joint function of both the hit rate, the probability of endorsing an old probe item as old, and the false alarm rate, the probability of incorrectly endorsing a new probe item as old. Neither the hit rate nor the false alarm rate by themselves place meaningful constraints on models of recognition memory. For instance, a high hit rate coupled with a high false alarm rate does not indicate good memory but rather a subject with a very liberal response bias. Only the relationship between the hit rate and false alarm rate contains information about discriminability. A single measurement of both hit rate and false alarm rate provides a single point in receiver operating characteristic (ROC) space.

ROC curves, constructed from observing memory at several different levels of confidence, combine several hit rate and false alarm rate pairs to trace out a curve in ROC space. Suppose we have an experiment in which the subject rates each recognition probe on a scale from one to six, with six corresponding to the highest rating of confidence that the probe item was presented on the list. We can calculate a meaningful hit rate and false alarm rate for each of five criteria. For instance, we might calculate the hit rate counting only "six" responses as endorsing the item as old. Alternatively, we could calculate hit rate counting both "five" and "six" responses as an endorsing the probe item as old. By calculating hit rate and false alarm rate for each possible criterion, we obtain a trajectory through ROC space. Because ROC curves measure discriminability simultaneously at several response criteria, they place a strong constraint on models of recognition performance.

The Yonelinas High Threshold Model (YHT).

The Yonelinas high threshold model (YHT) proposes that old test probes are recollected with probability R. If an old item is not recollected, the recognition decision relies on a familiarity process. Familiarity in the YHT is modeled as an equal variance signal-detection process. The strength of old probe items are drawn from a normal distribution. The strength of new probe items are drawn from a normal distribution with the same standard deviation as the old item distribution. The mean of the old item distribution is assumed to different from that of the new item distribution, a difference that is measured by a parameter d'_{YHT}, which measures the distance between the old and new item familiarity distributions in units of their common standard deviation. Although they are not necessarily strong assumptions of the model, in previously published work the probability of recollecting new items has been assumed to be zero and the standard deviations of old and new item familiarity distributions have been assumed to be equal.

The two components of the YHT give rise to distinctive patterns of responses in the ROC curve. Imagine for a moment that familiarity is absent from the recognition decision ($d'_{\mathrm{YHT}} = 0$) and the subject is relying solely on recollection. At the most conservative response criterion, the hit rate is R and the false alarm rate is zero. At the most liberal response criterion, the hit rate and false alarm rate are both one. As the response bias changes between conservative and liberal, the false alarm rate increases from zero based on guessing. Because there is no discriminability between old and new items on the basis of familiarity, the change in the hit rate associated with guessing is $1-R$ the change in the false alarm rate, resulting in a straight line. The YHT generates very different ROC curves if recollection is set to zero and d'_{YHT} is non-zero. Under these circumstances, a smooth ROC curve symmetric around the cross-diagonal results. Combining these two components, the result is a curved, asymmetric ROC curve that terminates on the hit rate axis.

Two processes or one, variable process?

In contrast to the view that two processes, recollection and familiarity, support recognition performance, a widespread view that is especially broadly held in the mathematical modeling community is that recognition performance is supported by a single, but variable process. One possibility is that the difference between old and new distributions is not limited to a change in the mean, as in the equal variance signal detection model, but is also associated with a change in the variability of the two distributions. This model is referred to as the normal unequal variance (NUV) signal detection model. The NUV would obtain, for instance, if the increment in strength each item acquired as a consequence of study was distributed normally. The NUV can be parameterized by d'_{NUV}, the difference between the means of the distributions, in units of the standard deviation of the new item distribution, and σ_O, the standard deviations of the old item distribution in units of the standard deviation of the new item distribution. Unlike the equal variance signal detection model, the NUV is able to generate asymmetric ROC curves consistent with those observed in typical recognition memory experiments. Although the YHT and NUV generate ROC curves that look similar to the naked eye, they make qualitatively different predictions about the form of the ROC curve when it is z-transformed.

In addition to the constraints provided by ROC curves, the z-transformed ROC curve, or zROC allows additional insight into the processes giving rise to recognition discriminability. If the recognition decision is the result of the NUV, then the zROC curve will appear linear, with the intercept of the zROC curve given by d'_{NUV} and the slope of the line given by σ_O. In contrast, the YHT predicts curvilinear zROC curves, with the curve inflected upward on the left hand side corresponding to the most conservative response criteria to the extent that $R > 0$. It should be noted that the YHT can predict linear zROC curves with a slope of 1 if $R = 0$. On this count, the bulk of the evidence lends support to the NUV; linear, or nearly linear, zROC curves with a slope significantly less than one are typically observed in item recognition (e.g., Glanzer, Kim, Hilford, & Adams, 1999; Heathcote, 2003; Hirshman & Hostetter, 2000; Ratcliff, Sheu, & Gronlund, 1992; Ratcliff, McKoon, & Tindall, 1994; Van Zandt, 2000) although not all studies report this finding (Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996). The widespread belief that zROC curves are linear for item recognition has led to models of item recognition that are designed specifically to generate linear zROC curves (e.g., Shiffrin & Steyvers, 1997; Chappell & Humphreys, 1994; McClelland & Chappell, 1998).

Although the YHT and NUV are good models to compare to each other—they have the same number of parameters yet embody very different assumptions about the processes underlying item recognition—they are not the only possible models of recognition performance. For instance, in order to explain data from subjects administered the anticholinergic scopolamine, Sherman, Atri, Hasselmo, Stern, and Howard (2003) proposed a variable recollection model, in which recollection is not an all-or-none process but can give rise to different gradations of confidence. Similar ideas have been proposed in dealing with data from associative recognition (Kelley & Wixted, 2001; Macho, 2004). In addition to the additional layer of complexity that can arise from considering more elaborate recollection, it is also possible to generate discrete-state models that abandon the concept of strength as a primitive concept entirely, yet produce a wide variety of ROC curves (Malmberg, 2002).

Previous work modeling aging using ROC curves

Given the theoretical constraints offered by ROC curves and the desirability of quantitative modeling of task performance in estimating the strength of cognitive processes, it is perhaps surprising that relatively little work has been done in fitting ROC curves to data from younger and older adults. Healy, Light, and Chung (2005) examined performance in an associative recognition task. Pairs of words were presented at study. At test, completely new pairs (e.g., study A-B, C-D, test E-F) as well as pairs composed of mismatched study items (e.g., study A-B, C-D, test A-D) were presented during a recognition test. Participants were instructed to reject both the new and re-paired lures. In addition to finding an age-related deficit in associative recognition, Healy et al. (2005) also found impaired item recognition for older adults comparing hit rate from intact pairs to false alarm rates from new pairs. They modeled item (intact *vs* new) and associative (intact *vs* rearranged) ROC curves using a variety of two-process models. Healy et al. (2005) found that the models of associative recognition converged in predicting that older adults had impaired recollection relative to younger adults. In contrast, however, the models of associative recognition produced divergent conclusions as to whether there is an age effect on familiarity. Complicating matters however, the model fits to the data from item recognition of pairs resulted in the opposite conclusion, that aging is associated with a decrement in familiarity, but did not have a significant effect on recollection.

Parks et al. (2004) examined ROC curves for younger and older adults performing item recognition for words presented auditorially in either the right or left ear in either a male or female voice. In the hard condition, participants were instructed to answer "recollect" in response to a test probe if they remembered in which ear the word was presented. In the vague condition, participants were instructed to answer "recollect" if they remembered any of the details of the probe item's presentation. If the probe wasn't recollected, Parks et al. (2004) collected a confidence rating on a six-point scale. By collapsing the recollect responses with the highest-confidence yes responses, Parks et al. (2004) were able to generate ROC curves for younger and older adults. In fitting the Yonelinas model to these ROC curves, Parks et al. (2004) found that both R and d'_{YHT} were lower for older adults than for younger adults, suggesting that both recollection and familiarity were affected by age. A potential limitation of this approach is that the YHT may not have provided a better fit to the experimental data than the NUV model, rendering the interpretation of the parameter estimates ambiguous. This concern is further exacerbated by the finding of linear zROC curves, a qualitative prediction of the NUV that is inconsistent with the YHT with non-zero recollection.

The present study differs from these previous efforts in a number of dimensions. We will study simple item recognition of travel pictures. This procedure should eliminate any potentially complicating factors associated with examining item recognition of pairs within the context of associative recognition[1] (Healy et al., 2005) or the effect of criterial recollective instructions (Parks et al., 2004). The use of travel pictures as to-be-remembered stimuli is a methodological difference compared to the bulk of previous recognition studies that have traditionally studied recognition for words. Previous studies have suggested robust recollection of travel pictures (Sherman et al., 2003; Schwartz, Howard, Jing, & Kahana, 2005). Finally, insofar as the YHT is quite controversial within the verbal learning community (Heathcote, 2003), we will fit both the YHT and the NUV to the data and evaluate the resulting fits before interpreting age differences in the parameters.

Experiment

In order to assess age differences in recollection and familiarity, we undertook a multiple-criterion item recognition experiment with younger and older adults. We examined both ROC curves and zROC curves to characterize the discriminability of younger and older adults at different response criteria. Subsequently we fit the YHT and NUV to the response distributions of each subject to evaluate the models as well as estimate recollection and familiarity across age groups.

Method

Participants. A total of 43 young adults and 33 older adults participated. Young adults were recruited from the Syracuse University community over the summer and consisted of a combination of undergraduates and graduate students. Older adults were recruited through the registry of the Adult Cognition Laboratory at Syracuse University. Mean age for the young adult group was 24.4 (SD = 2.7) and mean age for the older group was 71.2 (SD = 4.2). Women compromised 19/43 of the younger participants and 20/33 of the older participants. Mean years of education were 18.3 (SD = 2.6) for the young adults and 15.9 (SD = 3.0) for the older adults. Older adults performed a battery of standard cognitive tests (forward and backward digit span, identical pictures task, symbol digit task) on their first visit to the lab. Thirty-one of the older participants had previously served as a participant in another cognitive experiment in the Adult Cognition Laboratory, typically a skill learning experiment. Younger adults did not perform a battery of cognitive tests and were not previously participants in cognitive experiments conducted in our laboratories.

Procedure. Participants were given a picture recognition task in two sessions conducted at roughly the same time of day on two consecutive days. On each session, participants studied three lists of 128 digital pixmaps with resolution 350×232. The images subtended roughly 5 degrees of visual angle. Images were obtained from planetware.com, a travel picture website. The picture pool we used (Schwartz et al., 2005) includes a variety of scenes, outdoor and indoor, from travel destinations throughout the world, including nature pictures as well as urban scenes. The pool was constructed such that all pictures that contained text or that contained images that would be obviously emotionally salient to a large proportion of viewers (e.g. images of the World Trade Center in New York City) were eliminated from the pool. During study of the list, images were presented on the screen for 1 s, with a screen blank for 0.5 s between pictures. After each presentation of a list, participants were given 256 probes, half of which were in the list and half of which were new pictures. The test lists were constructed using an algorithm that attempted to equalize the distribution of relative lags, the difference in presentation serial position between adjacent old test items, such that relative lags with absolute value 1-5 were presented approximately equally often. This algorithm first assigned the old/new status of each test item randomly, and then attempted to insert relative lag pairs into pairs of successive old tests subjects to the constraint that no pictures were tested more than once. This algorithm was the same as that used in Schwartz et al. (2005). The results of the relative lag analyses will be described elsewhere.

In response to each test picture, subjects pressed a button from 1-6 to describe their confidence that the item was presented during study, with a 6 corresponding to an absolutely certain ”old” response and 1 corresponding to an absolutely “new” response. These responses were collected using a computer keyboard with the keys 1-3 replacing ‘z’ ‘x’ and ‘c’ and the keys 4-6 replacing ‘,’ ‘.’ and ‘/’ (standard qwerty layout). This allowed participants to respond comfortably using only the first three fingers of each hand. Participants were instructed to use all six buttons and to respond as quickly as possible without sacrificing accuracy. To familiarize participants with these procedures, participants studied and were tested on a practice list prior to administering the first study list. After the practice session, subjects were provided with feedback about their mean reaction time and the distribution of their responses. Subjects were not given explicit instructions as to what strategy they should use to encode the pictures, but were simply instructed to try to remember the pictures for a subsequent memory test.

The second experimental session was conducted on the day immediately after session 1. The two sessions were identical in procedure except that the participants completed a consent form and demographic measures at the beginning of Session 1. Over the two sessions, each subject responded to a total of 768 old item test probes and 768 new item test probes.

[1] For instance, it is possible that when studying pairs subjects search for relationships among the words. Strategic encoding and subsequent retrieval of this associative information could be used to reject new as well as rearranged lures. Any age differences in recollection observed in item recognition embedded in an associative paradigm could conceivably be attributable to older adults differentially adopting this strategy. Insofar as the item recognition paradigm places less emphasis on associative information, an item recognition experiment might be less susceptible to this type of criticism.

Modeling.

To assess the degree to which recollection and familiarity contributed to recognition performance for younger and older adults we fit the Yonelinas High Threshold model (YHT, Yonelinas, 1994, 1997, 2001) to each subject's response distribution. To do this we did a comprehensive search of R and d'_{YHT}. For each value of R and d'_{YHT} we constructed a model-derived ROC curve. We then slid the response criteria along the ROC curve incrementally to find the lowest possible χ^2 between the observed response distribution and the model's predictions. As far as we are aware, this precise technique has not been used previously. The data contains ten degrees of freedom corresponding to the 6 responses to old items and the 6 responses to new items. The model includes the two free parameters R and d'_{YHT} as well as the five response criteria. The fit therefore has three degrees of freedom for each subject.

Because the YHT has been criticized as a description of item recognition, we used the same procedure to fit the NUV to each participant's data. The procedure was identical except that the ROC curves were generated by the NUV with parameters σ_O and d'_{NUV}. We were able to compare the fits from the YHT to the fits obtained from the NUV as well as provide an alternative way to assess the effects of aging on model parameters.

Results

We examined the performance of younger and older adults using a variety of methods. First we examine hit rate, false alarm rate and d'. Next, we examine ROC and zROC curves. Then we describe the results of modeling the detailed pattern of performance across response criteria using the YHT and NUV.

Hit rates and false alarm rates.

We evaluated hit rate and false alarm rate for each of the possible criterion. We label the criterion according to the lowest response that would count as an endorsement of the item as old for the purposes of calculating hit rate. For instance, a response of "4" would be counted as a yes response for criteria 1-4, but not for criterion 5 or criterion 6. The hit rates for younger and older adults, along with t statistics for the comparisons are shown in Table 1. There were no significant differences between hit rate for younger and older adults at any of the response criteria.

The results for the false alarm rate calculations are shown in Table 1. At the most stringent response criteria, older adults showed a significantly higher false alarm rate than younger adults. The comparison at criterion 5 also remained significant when evaluated with a Wilcoxon signed-ranks test $(W = 912)$, $p < .05$. From this we conclude that there is a tendency for older adults to have higher false alarm rates than younger adults, but this tendency is only observed at the most stringent criteria.

Accuracy.

Of course it is impossible to learn anything about recognition discriminability by examining hit rate or false alarm rate in isolation. Figure 1 shows averaged untransformed ROC curves for younger and older adults. This just plots the hit rate as a function of the false alarm rate; both hit rate and false alarm rate can be found in Table 1. In reading this figure, recognition discriminability is read off informally by noting the distance of a point from the diagonal. The first thing that can be noted from the figure is that it appears that older adults use a less wide range of response biases, which can be seen from the fact that the older adults' ROC curve, while having a similar overall level of discriminability to that of younger adults, is less widely spread. Although discriminability for the groups is very similar, there are slight differences. The ROC curves for younger and older adults cross over, with slightly better discriminability for younger adults at more conservative response criteria (to the left of the figure) but slightly better discriminability for older adults at more liberal response criteria (to the right of the figure).

It would obviously be desirable to evaluate this impression about the crossover of age on discriminability with response criterion in a completely unbiased and theory-neutral way. Unfortunately, all possible measures of discriminability are based on some model. The most widely-used model of recognition discriminability is to calculate d' under the standard signal detection assumptions as $z(\text{HR}) - z(\text{FAR})$. This is equivalent to either the YHT with $R = 0$ or the NUV with $\sigma_O = 1$. Although this model of discriminability is certainly incorrect for these data (as can be seen clearly from the asymmetry of the observed ROC curves in Figure 1), d' is a widely-used measure of recognition discriminability. Table 1 shows d' for younger and older adults for each of the response criteria. At conservative criteria, younger adults show a tendency towards greater discriminability. This tendency diminished and even reversed, such that at more liberal response criteria older adults actually showed slightly higher discriminability than younger adults. Although none of the pairwise comparisons at the various response criteria showed a significant difference, the change in age differences appeared systematic with response criteria. To test for this, we conducted a repeated measures ANOVA on d' with age and criterion as factors. We found significant effects of response criterion, $F(4,296) = 116.4$, MSe $= 2.62$, $p < .001$, as well as a significant interaction of age and response criterion, $F(4,296) = 8.34$, MSe $= .19$, $p < .001$.

Although there were apparently not gross age differences in overall recognition performance, there is a change in the effect of age across response criteria. The foregoing analyses indicate that what age differences there are are manifest in relatively subtle properties of the shape of the ROC curves. We undertook subsequent analyses, including formal modeling, to reveal these differences.

zROC curves.

The YHT and NUV make qualitatively different predictions regarding the shape of the ROC curve when it is z-transformed, the zROC curve. We examined zROC curves for younger and older adults. The results of a regression analysis showed that both younger and older adults had a significant quadratic trend. For younger adults, the best-fitting in-

Table 1
Mean hit rate, false alarm rate and d′ for younger and older adults for each of the possible criterion values. A criterion value of four indicates that responses of four or greater are counted as yes responses for the purposes of this analysis. Standard errors are in parentheses. Significant differences $p < .05$ are presented in bold face.

		Criterion				
		6	5	4	3	2
Hit rate	Younger	.42 (.03)	.54 (.02)	.64 (.02)	.73 (.02)	.85 (.02)
	Older	.46 (.03)	.59 (.03)	.67 (.03)	.74 (.03)	.85 (.03)
	$t(74)$	-0.80	-1.25	-0.80	-0.23	0.10
False alarm rate	Younger	**.08 (.01)**	**.17 (.01)**	.29 (.02)	.42 (.03)	.61 (.04)
	Older	**.13 (.02)**	**.23 (.02)**	.31 (.02)	.43 (.03)	.62 (.04)
	$t(74)$	**-2.10**	**-2.46**	-1.01	-0.15	-0.04
d'	Younger	1.40 (.08)	1.14 (.07)	.98 (.07)	.89 (.07)	.84 (.07)
	Older	1.21 (.08)	1.04 (.06)	.99 (.06)	.92 (.07)	.87 (.06)
	$t(74)$	1.71	1.02	-.14	-.31	-.36

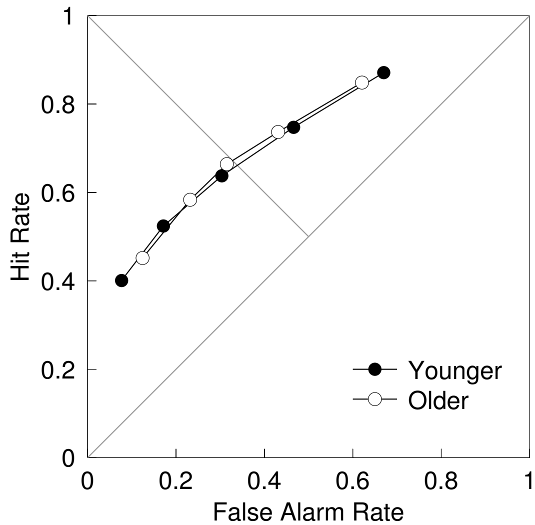

Figure 1. **Younger and older adults show comparable overall recognition accuracy.** Receiver operating characteristic (ROC) curves show hit rate as a function of false alarm rate for each of the five criteria that are possible with responses on a six-point scale. If there were no discrimination between old and new test probes, these curves would land on the diagonal running from the lower left to upper right corners. The extent to which the curves are above the diagonal reflects participants' ability to discriminate old from new items. ROC curves for younger and older adults overlap, suggesting comparable levels of overall recognition performance. Both curves show noticeable asymmetry with respect to the cross-diagonal. This rules out an equal variance signal detection process.

tercept was $0.83 \pm .06$ ($M \pm SE$), $t(42) = 13.4$, $p < .001$, the linear coefficient was $0.832 \pm .009$, $t(170) = 89.3$, $p < .001$ and the quadratic coefficient was $0.102 \pm .006$, $t(170) = 18.0$, $p < .001$. For older adults, the best-fitting intercept was $0.86 \pm .06$, $t(33) = 14.4$, $p < .001$, the linear coefficient was $0.87 \pm .01$, $t(134) = 68.6$, $p < .001$ and the quadratic coefficient was $0.084 \pm .009$, $t(134) = 9.1$, $p < .001$. The significant quadratic terms are inconsistent with the NUV, which predicts a linear zROC, but are consistent with the YHT.

Model fitting.

To properly compare differences in the recognition performance of younger and older adults it is necessary to first have a model that properly describes the shape of the ROC curve. Accordingly, we fit the YHT and the NUV to the ROC curves of each subject in the experiment. The YHT fit the young adult data better than did the NUV. The mean χ^2 for younger adults was 5.60 for YHT. The mean χ^2 for the NUV was 11.52, which was significantly higher, $t(42) = 4.30$, $p < .001$. The fit of the YHT to older adults' data was also somewhat better than that of the NUV. The average χ^2 for the fit of the YHT to older adults was 5.58, whereas the average χ^2 for the fit of the NUV was 7.90, $t(32) = 2.37$, $p < .03$.

Model parameters.

Armed with a model that describes the shape of the ROC curves, we can now ask about age differences in the model parameters. We examined model estimates of recollection and familiarity for younger and older adults. For the YHT, we found that the mean R for younger adults, $.30 \pm .02$ was significantly greater than the mean R for older adults $.23 \pm .02$, $t(72.9) = 2.22$, $p < .03$. In contrast, d'_{YHT} was

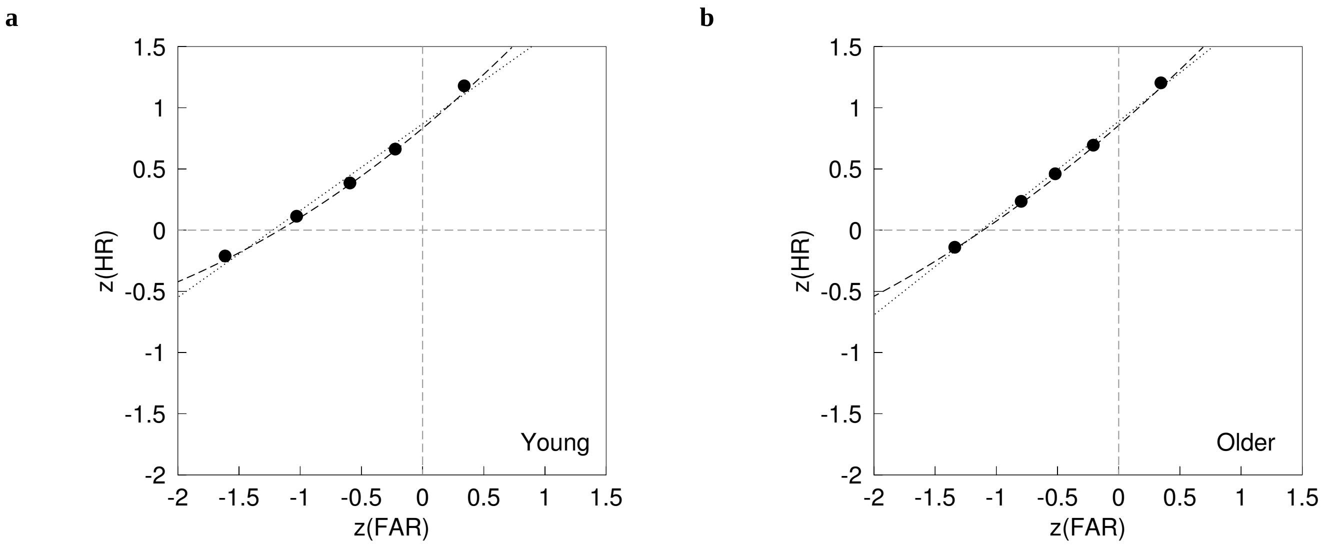

Figure 2. **z-transformed ROC curves for younger and older adults.** The results of linear (dotted) and quadratic (dashed) regressions are shown for each group. For each group, the quadratic coefficient was significant.

not greater for younger adults. Rather, there was a trend for the mean d'_{YHT} to be greater for older $.67 \pm .06$ than younger $.55 \pm .05$ adults, although this effect did not reach significance, $t(69.4) = 1.41$, $p < 0.15$. Figure 3 shows the best-fitting parameters from the YHT for each participant in this study.

Although the NUV model did not fit the data as well as the YHT model did, we nonetheless compared model parameters from the NUV across age groups. These provided convergent results for the conclusions we reached from the YHT. The variance of the old item distribution, σ_O was greater for young $1.46 \pm .04$ than for older $1.31 \pm .05$ adults, $t(64.9) = 2.50$, $p < .02$. In contrast, d'_{NUV} did not show an age effect. For younger adults the mean d'_{NUV} was $1.2 \pm .1$ whereas for older adults it was $1.09 \pm .08$. This difference was not significant, $t(73.5) = .92$. As for the YHT, the parameter of the NUV associated with skewed ROC curves, σ_O showed an age effect, whereas the parameter of the NUV corresponding to a univariate signal detection process, d'_{NUV}, did not show an age effect.

Given the fact that the YHT and NUV can produce identical ROC curves if $R = 0$ and $\sigma_O = 1$, it is natural to ask if the parameter estimates across models are correlated with each other. Collapsing across both subject groups, we observed significant correlations for both R and σ_O, $r = .67$, $t(74) = 7.69$, $p < .001$, and d'_{YHT} and d'_{NUV}, $r = .74$, $t(74) = 9.36$, $p < .001$. These correlations are a consequence of the strong correspondence between the model parameters.

Visual inspection of the best-fitting values of R and d'_{YHT} (Figure 3) suggested that model estimates of recollection and familiarity were not independent of each other across subjects. To assess this, we generated correlation coefficients between R and d'_{YHT} for younger and older adults. Interestingly, model estimates of R and d'_{YHT} were significantly correlated across participants for younger $r = 0.51$, $p < .001$, but not older $r = 0.15$, $p > .4$ participants. A similar pattern of results was found for the parameter estimates from the NUV in which σ_O and d'_{NUV} were significantly correlated across subjects for younger, $r = 0.58$, $p < .001$, but not older $r = 0.22$, $p > 0.2$ participants.

Figure 3 includes a regression line for younger adults. It appears that a cluster of older participants with high d'_{YHT} but very low R are responsible for the difference in the observed correlations in parameters across age groups.

General Discussion

We examined recognition memory of travel pictures in younger and older adults. For these groups of participants, there was no gross effect of age on overall recognition performance. Thus, any age differences would be attributable to relatively subtle changes in the pattern of responses across multiple criteria. The zROC curves for younger and older adults had significant quadratic curvature, suggesting that recognition memory was not well-described by the normal unequal variance (NUV) signal detection model. Consistent with the finding of nonlinear zROC curves, modeling results showed that the NUV model did not fit the response distributions for either younger or older adults as well as did the YHT model, which assumes a discrete recollective process in addition to a familiarity-based signal detection process. Despite the lack of a gross deficit in overall accuracy, older adults showed significantly lower levels of recollection than did younger adults. This finding confirms a number of other studies that argued that recollection, estimated in various ways in various experimental settings, is preferentially impaired in older adults (Bastin & Van Der Linden, 2003; Jennings & Jacoby, 1997; Quamme et al., 2004). In contrast, we found no evidence for an age deficit in familiarity. This is not consistent with other studies that have found that aging

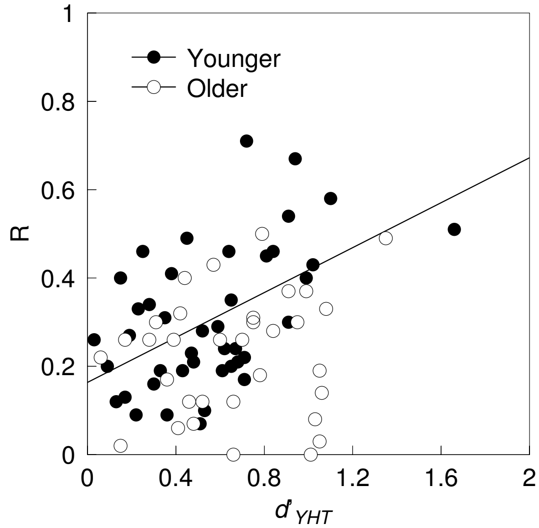

Figure 3. **Distribution of YHT model parameters from young and older subjects.** For each subject for whom YHT model parameters were available, R and d'_{YHT} are plotted as a function of each other. As a group, young subjects had a significantly higher average estimated R but the older adults had a non-significantly higher average d'_{YHT}. Interestingly, R and d'_{YHT} were significantly correlated with each other for young, but not older participants. The straight line is the result of a linear regression to the data from younger adults.

is associated with decreased familiarity (Parks et al., 2004; Prull et al., In press).

One discrepancy between the current study and the vast majority of previous work on the effects of aging on item recognition is the fact that we did not observe drastic effects of age on overall recognition performance. Although we observed some evidence for higher false alarm rates for older adults at the most stringent criteria (Table 1), the age differences we observed in modeling the data are best described as reflecting a relatively subtle change in the shape of ROC curves in older adults. There are several factors that might account for this. One possibility is that our use of travel pictures allowed older adults to make effective use of more extensive travel experience or perhaps more extensive geographical and cultural semantic knowledge in encoding the stimuli. Another possibility is that our older subject sample, the vast majority of whom had participated in previous cognitive experiments, was particularly motivated, or perhaps their prior experience with cognitive testing provided some very general practice effect. Further experimentation would clearly be necessary to distinguish these various possibilities. It is notable, however, that despite the lack of a dramatic effect of age on discriminability, there was nonetheless a reliable decrement in the proportion of recollective responses for older adults.

Recognition memory poses a number of challenges to the researcher interested in characterizing performance of different groups. In the present study, we were able to measure hit rate and false alarm rate for each of five criteria. Had we only collected yes-no responses we might have found either no deficit associated with aging, had the experiment worked out to have a liberal response criterion, or a deficit restricted to an increased false alarm rate for older adults had the experiment worked out to implement a conservative response criterion. By observing performance at multiple response criteria we were able to obtain a more complete picture of subjects' recognition memory. Moreover, there are considerable advantages to using cognitive models to describe the data. Traditional calculations of d' to characterize discriminability is essentially fitting a model to the data, in this case the equal variance signal detection model, a model that fails to describe observed skewed ROC curves. We were able to demonstrate that the YHT provides a fit superior to those of the equal variance signal detection and NUV models to the ROC curves of both younger and older subjects. The data in this experiment are composed of ten dependent variables. The use of the YHT enabled us to boil the pattern of responses across these ten dependent variables into a psychologically rich interpretation. Of course, this interpretation depends on the validity of the YHT, but this statement applies equally well to all models of recognition discriminability.

Interestingly, whereas estimates of recollection and familiarity were correlated across subjects for younger adults, no such relationship was observed for older adults. In the YHT, stochastic independence is typically assumed to hold between recollection and familiarity at retrieval. This greatly simplifies the equations that the model implements. The assumption of independence is also used in other experimental applications of two process theory, including the process dissociation procedure and corrected estimates of knowing in the remember/know procedure. While the finding of correlated estimates of recollection and familiarity in the present study does not bear on the question of stochastic independence at retrieval, it is worth pointing out that some recent criticisms of dual process theory (Dunn, 2004) are actually criticisms of the independence assumptions. Further, most authors argue not only that recollection and familiarity both depend on medial temporal lobe structures, but that the hippocampus supports recollection whereas surrounding regions of medial temporal lobe cortex support familiarity (e.g. Davachi, Mitchell, & Wagner, 2003; Ranganath et al., 2004; Fortin et al., 2004; Yonelinas et al., 2002). The hippocampus receives input from cortical regions of the MTL, so it is not hard to imagine that a strong form of stochastic independence between the processes would not hold. The reasons why aging might be associated in a weakening of the functional connection between recollection and familiarity are quite opaque at this time.

The Yonelinas high threshold model and the signal detection framework

This is the first study of which we are aware where the YHT and NUV models were directly compared to each other and the YHT was shown to generate a superior fit to the data. This was found separately for both young and old participants. Moreover, a qualitative prediction of the YHT, curvilinear zROC curves was observed, in contradiction of a qualitative prediction of the NUV. Although we are aware of other studies that have reported U-shaped zROC curves (Yonelinas et al., 1996; Sherman et al., 2003), neither of them have had nearly so much data per subject, nor so many subjects on which to report. Under at least some circumstances, the YHT provides an excellent account both of the distribution of subjects responses as well as the qualitative shape of zROC curves.

Given the widespread reports of linear (or nearly linear) zROC curves (Heathcote, 2003; Hirshman & Hostetter, 2000; Ratcliff et al., 1992, 1994; Van Zandt, 2000), what could account for the difference between the present results and earlier work showing that the NUV provides a better fit to ROC data (e.g., Heathcote, 2003)? One difference between the current study and previous work that showed linear zROC curves is the use of travel pictures as to-be-remembered stimuli. Although no formal analyses of a quadratic trend was conducted, Sherman et al. (2003) showed zROC curves that appeared curvilinear for their control subjects in a study of the effects of scopolamine on recognition performance. Secondary analyses of another study that used identical study materials as those used here with a slightly different procedure (Schwartz et al., 2005) also showed curvilinear zROC curves (M. W. Howard, unpublished observation). Travel pictures are not only perceptually detailed stimuli, but also undoubtedly provide a rich combination of semantic information as well. Moreover, the particular travel pictures used in this experiment are likely to be completely novel to the participants, whereas words are experienced in a wide variety of pre-experimental contexts. It is tempting to speculate that these properties of travel pictures make them particularly easy to recollect, leading to the strong explanatory power of the YHT in these studies.

If recognition performance for travel pictures is well-described by the YHT, but not the NUV, and recognition performance for words is well-described by the NUV, but not the YHT, this does not necessarily indicate qualitatively different memory process for different types of materials. One can describe the YHT in some sense as a traditional signal detection model (e.g., Norman & Wickelgren, 1969) in which the old item distribution is bimodal, with those items that are recollected having infinite strength. This assumption is untenable under sufficiently stringent response criteria. For instance, if subjects are offered a million dollars for every correct “no” response and no reward for every correct “yes” response, one can be reasonably well assured that subjects will never say yes to a probe, no matter how vividly they remember it. If recollection is not associated with infinite strength, then how best to characterize the strength of old items?

In order to describe the recognition performance of subjects administered scopolamine, Sherman et al. (2003) proposed a generalization of the YHT in which recollected items did not simply have infinite strength; rather, recollection was assumed to give rise to a normal distribution of strengths. That is, if an item is recollected, the resulting strength is variable, presumably reflecting different degrees of recollected detail. This variable recollection model, although not formally identical, is closely related to the some-or-none model proposed to describe associative recognition by Kelley and Wixted (2001). The some-or-none model has been explored by Macho (2004) and applied to the the effects of aging on associative recognition performance by Healy et al. (2005). One interesting feature of the variable recollection model is that it can generate zROC curves that are neither linear nor monotonically concave like those observed here. Imagine that the old item strength distribution is bimodal, with one distribution of strength corresponding to familiarity and another corresponding to recollection. For the purposes of illustration, assume further that the two peaks of the old item distribution are (very) widely spaced. As the ROC curve passes through the first (familiarity) peak, the ROC resembles that generated by the YHT. However, as the criterion becomes more conservative it starts to approach the second (recollective distribution). The ROC now moves towards the origin as the hit rate begins to decrease. When examined in zROC space, the phase where the criterion approaches the first distribution is linear, in the same way that the right side of the YHT’s zROC curve is linear. On the extreme left of the zROC curve, the zROC is also linear, but with a different slope. In between these two phases, there is a concave region, resulting in a “zig-zag” shape. If the criteria used in the experiment are not sufficiently conservative to enter the second linear region, the zROC curves correspond closely to those predicted by the YHT. The ability to describe zig-zag zROC curves was essential to describe the zROC curves obtained from participants administered scopolamine (Sherman et al., 2003).

The variable recollection framework includes the YHT as a special case and is sufficiently broad to be extremely difficult to distinguish from the NUV in practice. By turning down the variability on the recollective process in the variable recollection model, one recovers the YHT. When the difference between the recollected item distribution and the old familiarity-based distribution is small relative to the variability of the recollective distribution, the effect is similar to one broad old item distribution, as would be obtained from the NUV. Perhaps the variable recollection model or the some-or-none model offer a means to simultaneously account for the nearly-linear zROC curves typically obtained with item recognition of words and the curvilinear zROC curves that have been observed here with travel pictures as stimuli.

Aging, recollection and the hippocampus

The present results provide support from the quantitative modeling of ROC curves from item recognition for a broad literature arguing that aging is associated with a preferential decrement in recollection (e.g., Jennings & Jacoby, 1997; Healy et al., 2005; Naveh-Benjamin, 2000; Naveh-Benjamin et al., 2003; Prull et al., In press). Recent work has shed light on the possible neural basis that underlies this deficit in recollection. Theorists have for some time hypothesized that the hippocampus is particularly important in the recollective component of recognition performance (e.g., Aggleton & Brown, 1999; Norman & O’Reilly, 2003). Recent evidence from neuroimaging implicates the hippocampus proper in recollection (Yonelinas, Otten, Shaw, & Rugg, 2005; Ranganath et al., 2004; Davachi et al., 2003) and episodic memory more generally (Addis, Moscovitch, Crawley, & McAndrews, 2004; Gilboa, Winocur, Grady, Hevenor, & Moscovitch, 2004). Fortin et al. (2004) showed that ROC curves from rats performing an item recognition task for odors were consistent with those predicted by the YHT. However, rats with lesions to the hippocampus generated ROC curves that were well-predicted by an equal variance signal detection model. There is therefore ample evidence that the hippocampus supports recollection. There is also ample evidence that normal aging is associated with a disruption of the hippocampal formation (see Rosenzweig & Barnes, 2003, for a recent review). Small, Chawla, Buonocore, Rapp, and Barnes (2004) argued on the basis of both functional neuroimaging and studies of gene expression that the changes in the hippocampal formation associated with aging were restricted to the dentate gyrus, a subregion of the hippocampal formation. Although aging is associated with physiological changes in the hippocampus and the hippocampus is implicated in recollection, the deficit observed in older adults' recollection is not necessarily a direct consequence of changes in the hippocampus proper. Older adults have well-documented changes in prefrontal regions (West, 1996; Head et al., 2004; Raz et al., 1997), and the frontal cortex and hippocampus may be closely linked physiologically (Hyman, Zilli, Paley, & Hasselmo, 2005; Siapas, Lubenov, & Wilson, 2005). Perhaps the changes in hippocampal function presumed to underlie the deficit in recollection for older adults are actually a consequence of changes in prefrontal functioning.

The present results also suggest that the recollective process observed in item recognition is analogous to the recovery of temporal context hypothesized to support episodic associations in recall (Howard & Kahana, 2002; Howard, Fotedar, Datey, & Hasselmo, 2005). Howard, Wingfield, and Kahana (In press) fit the temporal context model (TCM) to associative functions from free recall of random word lists of younger and older adults (Kahana et al., 2002). They found that the deficit in temporally-defined associations observed with aging was well-described by allowing the parameter controlling item-to-context binding to be lower for older adults. Howard et al. (2005) hypothesized that this parameter is a function of the hippocampus proper and used this hypothesis to describe neuropsychological findings (Bunsey & Eichenbaum, 1996). There is also evidence that recollection in item recognition is associated with recovery of temporal context. Schwartz et al. (2005) showed that when a test item is recollected, memory for following test items is enhanced if those test items were from nearby serial positions. Taken together, these data suggest that the mnemonic deficit in aging is associated with a decrease in the ability of the hippocampus to enable binding of items to the temporal-spatial context in which they were presented.

Conclusions

We studied item recognition of travel pictures in younger and older adults. We collected multiple confidence levels to enable the construction of ROC curves. Overall levels of accuracy were comparable for the younger and older adults used in this study. We found evidence for non-linear zROC curves in item recognition for both younger and older adults, suggesting than the NUV account of recognition performance did not describe the data. Indeed, the YHT provided a better fit to the observed response distributions. Estimates of recollection were significantly higher for younger than older adults. Estimates of familiarity were higher for older adults than younger adults, although the difference was not significant. This finding, coupled with other recent work, supports the hypothesis that the deficit in episodic memory with aging is a consequence of a failure to bind items to the temporal-spatial context in which they are experienced, a function that is probably dependent on the integrity of the hippocampus.

References

Addis, D. R., Moscovitch, M., Crawley, A. P., & McAndrews, M. P. (2004). Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. *Hippocampus*, *14*(6), 752-62.

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral Brain Science*, *22*(3), 425-44; discussion 444-89.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (p. 243-290). San Francisco: Freeman.

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.

Bunsey, M., & Eichenbaum, H. B. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, *379*(6562), 255-257.

Chappell, M., & Humphreys, M. (1994). An autoassociative neural network for sparse representations: Analysis and application to models of recognition and cued recall. *Psychological Review*, *101*, 103-128.

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Science, USA*, *100*(4), 2157-62.

Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychological Review*, *111*(2), 524-542.

Dywan, J., & Jacoby, L. L. (1990). Effects of aging on source monitoring: differences in susceptibility to false fame. *Psychology and Aging*, *5*(3), 379-87.

Fortin, N. J., Wright, S. P., & Eichenbaum, H. (2004). Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature*, *431*(7005), 188-91.

Gilboa, A., Winocur, G., Grady, C. L., Hevenor, S. J., & Moscovitch, M. (2004). Remembering our past: functional neuroanatomy of recollection of recent and very remote personal events. *Cerebral Cortex*, *14*(11), 1214-25.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *25*(2), 500-513.

Head, D., Buckner, R. L., Shimony, J. S., Williams, L. E., Akbudak, E., Conturo, T. E., McAvoy, M., Morris, J. C., & Snyder, A. Z. (2004). Differential vulnerability of anterior white matter in nondemented aging with minimal acceleration in dementia of the alzheimer type: evidence from diffusion tensor imaging. *Cerebral Cortex*, *14*(4), 410-23.

Healy, M. R., Light, L. L., & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *31*, 768-788.

Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *29*(6), 1210-30.

Hintzman, D. L., & Curran, T. (1994). Retieval dynamics of recognition and frequency judgements: Evidence for separate processes of famaliarity and recall. *Journal of Memory and Language*, *33*, 1-18.

Hirshman, E., & Hostetter, M. (2000). Using ROC curves to test models of recognition memory: the relationship between presentation duration and slope. *Memory & Cognition*, *28*(2), 161-6.

Howard, M. W., Fotedar, M. S., Datey, A. V., & Hasselmo, M. E. (2005). The temporal context model in spatial navigation and relational learning: Toward a common explanation of medial temporal lobe function across domains. *Psychological Review*, *112*(1), 75-116.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269-299.

Howard, M. W., Wingfield, A., & Kahana, M. J. (In press). Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. *Psychonomic Bulletin & Review*.

Hoyer, W. J., & Verhaeghen, P. (In press). Memory aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging (sixth edition).* San Diego: Elsevier.

Hyman, J. M., Zilli, E. A., Paley, A. M., & Hasselmo, M. E. (2005). Medial prefrontal cortex cells show dynamic modulation with the hippocampal theta rhythm dependent on behavior. *Hippocampus*, *15*(6), 739-749.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541.

Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology and Aging*, *12*(2), 352-61.

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag-recency effects in free recall. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *28*, 530-540.

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *27*(3), 701-22.

Light, L. L., Patterson, M. M., Chung, C., & Healy, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. *Memory & Cognition*, *32*(7), 1182-1193.

Macho, S. (2004). Modeling associative recognition: a comparison of two-high-threshold, two-high-threshold signal detection, and mixture distribution models. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *30*(1), 83-97.

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *28*(2), 380-7.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252-271.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: a subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*(4), 724-760.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: tests of an associative deficit hypothesis. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *26*(5), 1170-87.

Naveh-Benjamin, M., Guez, J., Kilb, A., & Reedy, S. (2004). The associative memory deficit of older adults: further support using face-name associations. *Psychology and Aging*, *19*(3), 541-6.

Naveh-Benjamin, M., Hussain, Z., Guez, J., & Bar-On, M. (2003). Adult age differences in episodic memory: further support for an associative-deficit hypothesis. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *29*(5), 826-37.

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in short-term memory. *Journal of Mathematical Psychology*, *6*, 192-208.

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), 611-46.

Onyper, S. V., Hoyer, W. J., & Cerella, J. (in press). Determinants of retrieval solutions during cognitive skill training: Source confusions. *Memory & Cognition*.

Parks, C. M., Toth, J. P., & Smith, A. D. (2004). Aging and noncriterial recollection in the remember-know and dual process signal detection procedures. *Poster presented at the 45th Annual Meeting of the Psychonomic Society*.

Prull, M. W., Crandell-Dawes, L. L., Martin, A. M., Rosenberg, H. F., & Light, L. L. (In press). Recollection and familiarity in recognition memory: Adult age differences and neuropsychological test correlates. *Psychology and Aging*.

Quamme, J. R., 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), 672-91.

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*(1), 2-13.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory ROC functions and implications for GMMs. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *20*, 763-785.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*(3), 518-535.

Raz, N., Gunning, F. M., Head, D., Dupuis, J. H., McQuain, J., Briggs, S. D., Loken, W. J., Thornton, A. E., & Acker, J. D. (1997). Selective aging of the human cerebral cortex observed in vivo: differential vulnerability of the prefrontal gray matter. *Cerebral Cortex*, *7*(3), 268-82.

Rosenzweig, E. S., & Barnes, C. A. (2003). Impact of aging on hippocampal function: plasticity, network dynamics, and cognition. *Progress in Neurobiology*, *69*(3), 143-179.

Rotello, C. M., & Heit, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory and Language*, *40*, 432-453.

Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. *Psychological Science*, *16*(11), 898-904.

Sherman, S. J., Atri, A., Hasselmo, M. E., Stern, C. E., & Howard, M. W. (2003). Scopolamine impairs human recognition memory: Data and modeling. *Behavioral Neuroscience*, *117*, 526-539.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM — retrieving effectively from memory. *Psychonomic Bulletin and Review*, *4*, 145-166.

Siapas, A. G., Lubenov, E. V., & Wilson, M. A. (2005). Prefrontal phase locking to hippocampal theta oscillations. *Neuron*, *46*(1), 141-51.

Small, S. A., Chawla, M. K., Buonocore, M., Rapp, P. R., & Barnes, C. A. (2004). Imaging correlates of brain function in monkeys and rats isolates a hippocampal subregion differentially vulnerable to aging. *Proceedings of the National Academy of Science, USA*, *101*(18), 7181-6.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1-12.

Van Zandt, T. (2000). ROC curves and confidence judgements in recognition memory. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *26*(3), 582-600.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*(2), 272-92.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, *11*(4), 616-41.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *20*(6), 1341-54.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, *25*, 747-763.

Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society B*, *356*(1413), 1363-74.

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

Yonelinas, A. P., Dobbins, I. G., Szymanski, M. D., Dhaliwal, H. S., & King, S. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition: An International Journal*, *5*, 418-441.

Yonelinas, A. P., Kroll, N. E., Quamme, J. R., Lazzara, M. M., Sauvé, M. J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience*, *5*(11), 1236-41.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, *25*(11), 3002-8.