

# Recall and recognition are equally impaired in patients with selective hippocampal damage

JOHN T. WIXTED and LARRY R. SQUIRE

*University of California at San Diego, La Jolla, California*

In two recent studies, the hypothesis that recall is more severely impaired than recognition in patients with damage thought to be limited to the hippocampus was tested. Yonelinas et al. (2002) reported findings that appeared to support this hypothesis, whereas Manns, Hopkins, Reed, Kitchener, and Squire (2003) found that recall and recognition were equally impaired. An analysis of the individual subject data from the two studies revealed that the apparent disagreement stemmed from the inclusion of a single outlying recognition score for 1 of the 55 control subjects in Yonelinas et al. (2002). When that outlier was excluded, the studies were in agreement that recognition and recall are substantially and similarly impaired in these patients. Yonelinas et al. (2002) also analyzed remember/know judgments and ROC data in an effort to show that recollection is selectively impaired in patients, but these analyses also raise problems.

Over the last 2 decades, a great deal of evidence has accumulated suggesting that human amnesia is associated with a selective impairment of memory, such that declarative memory is impaired and nondeclarative memory is preserved (Schacter & Tulving, 1994; Squire, 1992). A more controversial issue concerns whether or not particular aspects of declarative memory are differentially impaired in amnesia as well (Brown & Aggleton, 2001; Rugg & Yonelinas, 2003). Declarative memory is typically tested using either recall (the ability to reproduce a previously presented item) or recognition (the ability to identify an item as having been recently encountered). One view is that recall is based solely on a recollection process, whereas recognition is based on both recollection and familiarity (e.g., Mandler, 1980). The question of interest is whether recollection is more severely impaired in amnesia than is familiarity.

In their review of the literature, Yonelinas, Kroll, Dobbins, Lazzara, and Knight (1998) concluded that recollection and familiarity are both substantially impaired in amnesic patients with large left unilateral medial temporal lesions. However, in a more recent report, Yonelinas et al. (2002) proposed that a selective impairment in recollection can be observed in patients with damage limited to the hippocampus. According to their proposal, the hippocampus supports recollection, whereas the adjacent medial temporal lobe cortex that lies along the parahippocampal gyrus supports familiarity. If this account is correct, patients with damage limited to the hippocampus should exhibit a more pronounced impairment when declarative

memory is tested by recall (which is fully dependent on recollection) than when it is tested by recognition (which can be supported by familiarity when recollection fails).

Yonelinas et al. (2002) tested this prediction by comparing the recall and the recognition performance of 56 hypoxic patients (cardiac arrest survivors), using the Rey Auditory Verbal Learning Test (RAVLT). These patients were presumed to suffer from damage limited to the hippocampus (although no radiological or histological data were available to support that presumption) and were, therefore, expected to perform more poorly on the recall subtest of the RAVLT than on the recognition subtest. Because recall and recognition were measured on different scales, performance on the two tests was compared after first converting the data to  $z$  scores based on the recall and recognition performance of a large control group ( $n = 55$ ). The results were that the average recall  $z$  score for the hypoxic patients was significantly more negative than the average recognition  $z$  score (i.e., recall performance was more impaired than recognition performance). This finding is consistent with the idea that the hippocampus is more important for recollection than for familiarity.

Yet the results reported by Yonelinas et al. (2002) contrast with other findings in the literature (Haist, Shimamura, & Squire, 1992; Manns, Hopkins, Reed, Kitchener, & Squire, 2003). In particular, Manns et al. (2003) compared the recall and the recognition performance of patients with damage limited to the hippocampal region (the hippocampus proper, the dentate gyrus, and the subiculum), and they used the same memory test that Yonelinas et al. (2002) used (the RAVLT). The selective hippocampal lesions for 6 of the 7 patients in this study were confirmed by a quantitative analysis of magnetic resonance images (MRIs). In contrast to the findings reported by Yonelinas et al. (2002), Manns et al. (2003) found that recall and recognition were substantially and equivalently impaired. The contrast be-

---

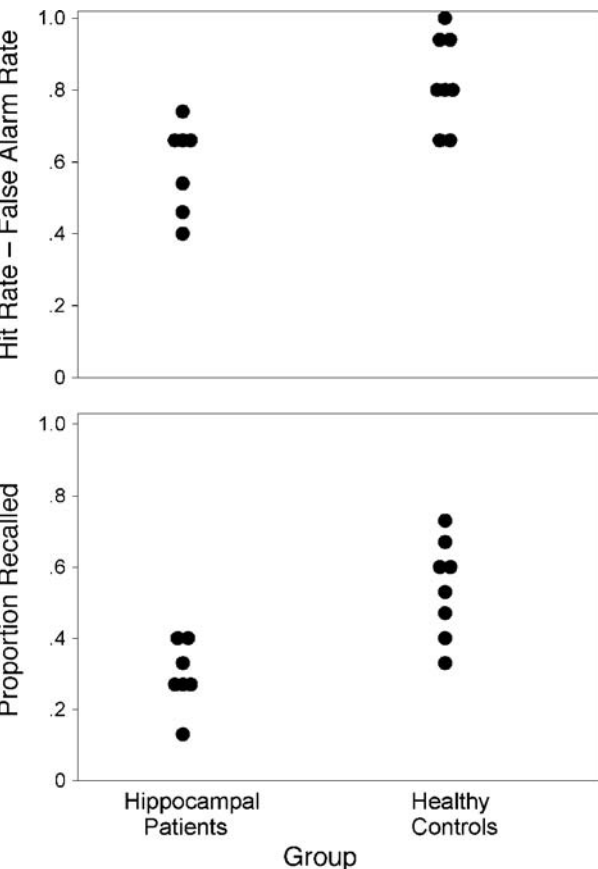
Correspondence concerning this article should be addressed to either J. T. Wixted or L. R. Squire, Department of Psychology, University of California at San Diego, La Jolla, CA 92093 (e-mail: jwixted@ucsd.edu or lsquire@ucsd.edu).

tween these two studies is striking given their apparent similarities, and the primary purpose of the present article is to examine the relevant data in some detail in an effort to reconcile the opposing results.

Our analyses will focus mainly on the first experiment reported by Yonelinas et al. (2002). That experiment involved large patient and control groups (56 and 55 subjects, respectively) and, thus, might be expected to have high power to detect any differences in recall and recognition that might exist. After reviewing those data in relation to the similar experiment performed by Manns et al. (2003), we also will briefly consider the results of two much smaller experiments (involving groups of 3 or 4 patients) that were reported by Yonelinas et al. (2002).

### Recall and Recognition Reanalysis

Neither Yonelinas et al. (2002) nor Manns et al. (2003) reported the distributions of raw data from the individual subjects, which is not unusual, but those data reveal why these studies came to such different conclusions. Figure 1 presents the distribution of recall (percentage correct) and recognition (hit rate minus false alarm rate) scores for the



**Figure 1.** Individual recall (proportion recalled) and recognition scores (hit rate minus false alarm rate) for hippocampal patients ( $n = 7$ ) and healthy controls ( $n = 8$ ) from Manns, Hopkins, Reed, Kitchener, and Squire (2003).

7 amnesic patients and 8 controls from Manns et al. (2003). The data are taken from the first of five recall and recognition trials,<sup>1</sup> and each symbol represents the score of an individual subject. These distributions are unremarkable in every respect. That is, visual inspection reveals no obvious outliers, the distributions are essentially symmetric about their means, and the data show no evidence of floor or ceiling effects. The recognition scores for the controls approached the ceiling, but only 1 subject achieved the maximum score of 1.0. Moreover, the highest and lowest recognition scores obtained by the controls are approximately equidistant from the mean score of .83, so there is no suggestion that the distribution is substantially pressed against the ceiling (i.e., the upper end of the distribution does not seem to be truncated by a ceiling effect).

Given the unremarkable distributions of recall and recognition scores for the controls, it makes sense to convert the corresponding scores for the amnesic patients to  $z$  scores, according to the following formula:

$$z_i = (s_i - \mu) / \sigma,$$

where  $z_i$  is the recognition (or recall)  $z$  score for a particular patient,  $s_i$  is the raw recognition (or recall) score for that patient,  $\mu$  is the corresponding mean score of the control group, and  $\sigma$  is the standard deviation of those scores. Because the raw scores for the amnesic patients fall below the mean score of the controls, all of the amnesic  $z$  scores are negative. The mean amnesic  $z$  scores for recall and recognition in Manns et al.'s (2003) study were  $-1.83$  and  $-1.91$ , respectively. These scores reflect a considerable degree of impairment, but the degree of impairment does not differ depending on whether declarative memory is tested by recall or recognition.

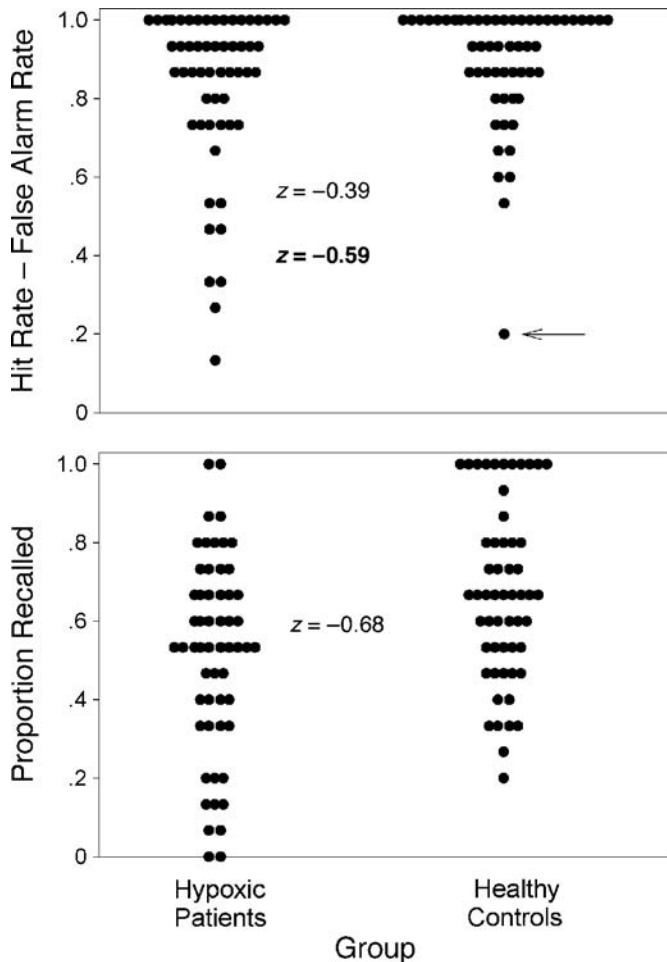
From the  $z$  score formula presented above, it is apparent that an inflated estimate of  $\sigma$  will reduce the  $z$  score estimate for every one of the amnesic scores. This fact is not particularly relevant to Manns's et al. (2003) data (because there is no reason to suppose that the estimate of  $\sigma$  is inaccurate), but this issue looms large in the data set collected by Yonelinas et al. (2002). The raw data from that study were kindly provided to us by the authors, and Figure 2 presents the individual recall and recognition scores for both the hypoxic patients and the controls. The data are taken from a delayed memory test conducted 20 min after the subjects had completed five recall and recognition study-test trials (the data from the first of those five trials, which would correspond exactly to the data published by Manns et al. 2003, were not reported). Visual inspection of these data reveals obvious ceiling effects, as well as an outlier (indicated by the arrow) in the recognition scores for the controls. The outlying score is so low that only 1 of the 56 hypoxic patients obtained a lower recognition score. Moreover, the outlying score falls more than 2 standard deviations below the next lowest recognition score and more than 4.4 standard deviations below the mean recognition score of the controls. No other score for the control subjects in either the recall distribution or the recognition distribution is more than 0.5 standard devia-

tions removed from an adjacent score or more than 2.3 standard deviations below the mean.

Although no method exists to identify definitively a non-representative outlier, Tukey (1977) offered a useful “rule-of-thumb” method in his book on exploratory data analysis. The method involves computing the 25%, 50%, and 75% quartiles (Q1, Q2, and Q3), as well as the interquartile range (IRQ, which equals Q3 – Q1), and then establishing “fences” at various points beyond that range. The “lower inner fence,” for example, is defined to be Q3 – 1.5IRQ, whereas the “lower outer fence” is defined to be Q3 – 3.0IRQ. As a rule of thumb, values that fall between the inner and the outer fences are regarded as possible outliers, and values that fall beyond the outer fence are regarded as extreme outliers (Tukey, 1977). With this method, only one score is identified as an extreme outlier in all of

the distributions in Figures 1 and 2, and that score is the one that stands out visually in Figure 2.

With the outlier included, the standard deviation of recognition scores for the controls is relatively large, and the mean recognition score is lower than it otherwise would be. As a result, the recognition z score deficit for the hypoxic patients is small, and the mean recognition z score (–0.39) is much lower than the mean recall z score [–0.68;  $t(55) = 2.26, p = .027$ ]. However, with the outlier excluded, the standard deviation of recognition scores for the controls becomes noticeably smaller, the mean score becomes larger, and the difference between the amnesic recall and recognition z scores (now –0.68 and –0.59, respectively) is negligible [ $t(55) = 0.62, p = .54$ ]. Although a small differential deficit remains numerically apparent (–0.68 vs. –0.59), one would suppose that the high power



**Figure 2. Individual recall and recognition scores for hypoxic patients ( $n = 55$ ) and healthy controls ( $n = 56$ ) from Yonelinas et al. (2002). The arrow indicates an outlier in the control recognition scores. The z scores are the mean scores for the patients, suggesting that recall was impaired more than recognition. This difference disappears (see the recognition z score in bold) when the outlier is excluded.**

afforded by such large sample sizes would have rendered that difference statistically significant if the effect were real. Then again, the data are seriously compromised by the presence of ceiling effects, which may have diminished to some extent the benefits of a large sample size.

The patients studied by Manns et al. (2003) had mean recall and recognition  $z$  scores of  $-1.83$  and  $-1.91$ , respectively, whereas those studied by Yonelinas et al. (2002) had mean recall and recognition  $z$  scores of  $-0.68$  versus  $-0.59$ , respectively (with the outlier removed). This difference in overall severity of memory impairment, which may be attributable to more extensive hippocampal damage in the patients tested by Manns et al. (2003), does not appear to be relevant to the question of whether or not recall is more severely impaired than recognition. With the outlier excluded, even the 10 hypoxic patients with the most severely impaired recall scores in Yonelinas et al.'s (2002) study—those that fall more than 2 standard deviations below the mean of the control subjects—exhibit no evidence of a differential deficit (recall  $z$  score =  $-2.48$ , recognition  $z$  score =  $-2.55$ ).

We conclude from this analysis that the most likely reason for the discrepancy between the results reported by Manns et al. (2003) and Yonelinas et al. (2002) is that the latter study included a deviant score from a single control subject that exerted undue influence on the analysis of the recall and recognition performance of the patients. When that deviant score is excluded, the results from the two studies are quite similar, and they suggest that, regardless of whether memory is tested by recall or recognition, declarative memory is impaired to an equivalent degree in patients with damage limited to the hippocampal region. Thus, if recognition memory is supported by a familiarity process and a recollection process, both processes would seem to be dependent on an intact hippocampus.

Yonelinas et al. (2002) conducted additional analyses on the recall and recognition  $z$  scores, which were also taken to support the idea that recall is selectively impaired in hippocampal patients. Their Figure 1b, for example, presented a scatterplot of the recall  $z$  scores versus the recognition  $z$  scores of the hypoxic patients. Superimposed on those data were the results of a linear regression performed on the recall and recognition  $z$  scores of the control subjects. This plot suggested that most of the patients had a more pronounced recall than recognition deficit, because most of the patient scores fell below the regression line of the control subjects. That is, in relation to the controls, recall scores were lower than they should have been for a given level of recognition performance. As will be described below, this finding is not noticeably influenced by the inclusion or exclusion of the outlier, so it is worth considering whether or not it points to a selective recall deficit in the patients after all.

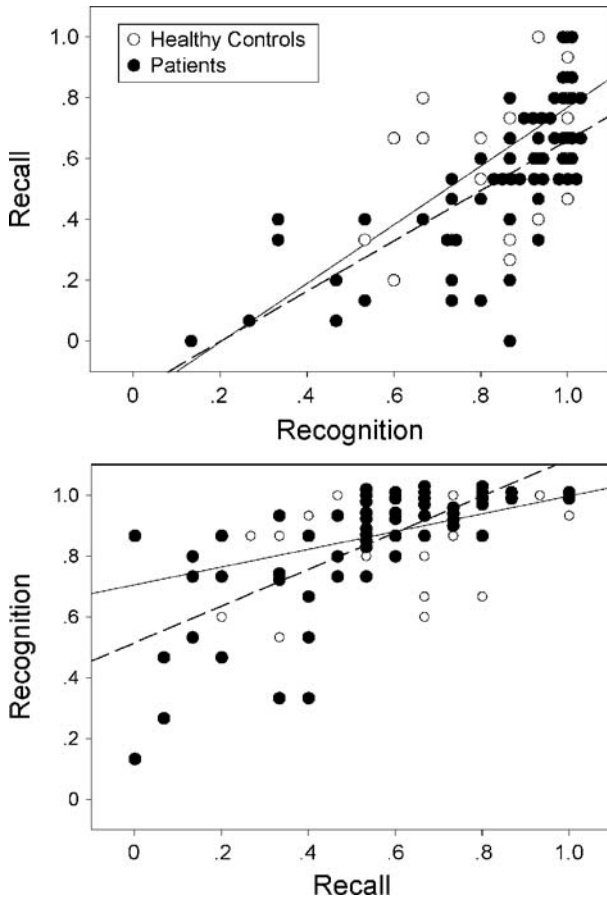
The regression analysis could, in principle, be performed on the raw recall and recognition scores or on the  $z$ -transformed versions of those scores. The conversion to  $z$  scores is simply a linear transformation of the data, which will not affect the correlation between the recall and the recognition measures or the relative slopes of the regres-

sion lines for patients and controls. That is, if the slope of the patient regression line is half that of the control regression line when raw scores are analyzed, the same will hold true when the corresponding  $z$  scores are analyzed. However, because the zero point of the  $x$ -axis changes when the data are converted to  $z$  scores, and because the intercept of the regression line is the estimated value of  $y$  when  $x$  equals zero, the intercept will change depending on whether raw scores or  $z$  scores are analyzed. With respect to the raw scores, the zero point on the  $x$ -axis represents the lowest possible level of performance. Thus, if recall is plotted on the  $x$ -axis, the intercept of the regression line is to the left of the plot and reflects the predicted level of recognition performance when the proportion of items recalled is zero. By contrast, with respect to  $z$  scores, the zero point on the  $x$ -axis represents the mean level of performance for the controls. If recall is plotted on the  $x$ -axis, the intercept reflects the predicted level of recognition performance (in  $z$  score units) when the proportion of items recalled is equal to the mean recall score for the controls. In this case, the intercept is not at the left of the plot.

It is arbitrary whether the regression analysis is performed on recall plotted against recognition (as Yonelinas et al., 2002, did) or on recognition plotted against recall. Accordingly, we reanalyzed both the raw scores and the  $z$  scores and plotted recall as a function of recognition, as well as recognition as a function of recall. Even with the outlier removed, the validity of these regression analyses is open to question because of the ceiling effects evident in the data. Nevertheless, we carried out these analyses because Yonelinas et al. (2002) reported a similar analysis, and it is important to determine whether the small non-significant difference in the recall and recognition scores of the patients can be rendered significant when the data are analyzed in other ways.

The upper panel of Figure 3 shows the raw recall scores from Yonelinas et al. (2002) plotted against the raw recognition scores for the patients (filled symbols) and controls (open symbols), as well as the regression line for each (solid line for the controls, dashed line for the patients). The question of interest is whether the regression line for the patients differs from that of the controls, and the answer, in this case, is that it does not. Neither the slope nor the intercept of the regression line that relates recall and recognition scores for the patients differs significantly from the corresponding parameters of the regression line that relates recall and recognition scores for the controls [ $t_s(108) < 0.6$ ].

The lower panel of Figure 3 shows the same data, except that now the dependent measures are plotted in opposite fashion (i.e., recognition is plotted against recall). Now the parameters of the regression lines for the patients and the controls do differ significantly. Specifically, the regression line for the patients exhibits a steeper slope and a smaller intercept [ $t_s(108) > 2.80$ ] than does the regression line for the corresponding values for the controls. The intercept in this case is the estimated recognition score when recall is at 0. Surprisingly, this outcome would appear to suggest that in relation to controls, recognition performance was poorer (not better) than it should be for a



**Figure 3.** Recall versus recognition raw scores (upper panel) and recognition versus recall raw scores (lower panel) for patients (filled symbols) and healthy controls (open symbols) (data from Yonelinas et al., 2002, with the outlier excluded). Some symbols represent data for more than 1 individual who received the same recall and recognition scores. The solid line represents the regression line for the controls, and the dashed line represents the regression line for the patients.

given level of recall performance, at least when recall is severely impaired.

The upper panel of Figure 4 shows the same analysis as that shown by the upper panel of Figure 3, after converting the data to  $z$  scores. Even though the outlier has been excluded, this plot appears similar to that in Figure 1b of Yonelinas et al. (2002). The zero point on the  $x$ -axis now represents the average level of recognition performance for the controls, and the intercept estimate is the estimated level of  $z$  recall at that point (not at the left of the plot, where  $z$  recognition =  $-8$ ). Although the slopes of the regression lines do not differ significantly, the intercept of the patient regression line is significantly lower than that of the controls [ $t(108) = 2.94$ ]. This result is consistent with the idea that recall is more impaired than recognition, at least for patients with an average level of recognition performance, relative to the controls. However, when the same data are plotted as  $z$  recognition versus  $z$  recall, instead

of  $z$  recall versus  $z$  recognition (lower panel, Figure 4), the story changes. Whereas one might now expect to see recognition scores that are too high relative to the recall scores, no such effect emerged. The slope of the regression line for the patients is steeper than that of the controls (which must be true, because it was true of the raw scores), but the intercepts at  $z$  recall = 0 do not differ [ $t(108) < 0.80$ ].

The fact that the conclusion differs depending on how the data are plotted (recall vs. recognition or recognition vs. recall) is probably attributable to the fact that several assumptions underlying the regression test are violated in these analyses. For example, regression tests assume that only the  $y$  variable is associated with measurement error, whereas in this case both the  $x$  and the  $y$  variables are associated with measurement error (because they are both dependent measures). Also, regression tests assume that error variance in the  $y$  measure is normally distributed, but this cannot be true given the evident ceiling effects. Thus, although the hint of a differential recall deficit is apparent in one of the four possible regression analyses (the one illustrated by Yonelinas et al., 2002, and illustrated without the outlier in Figure 4, top panel), most of the analyses suggest that recall and recognition were similarly impaired in the hypoxic patients. This is the same conclusion as is reached by simply comparing the overall deficit in recall and recognition, as was described earlier.

It is also useful to note that additional analyses reported by Yonelinas et al. (2002) in their Figures 2a and 2b also suffer from ceiling effects and the effects of an outlier. Furthermore, these analyses involve theoretical assumptions derived from a specific quantitative model of recognition memory advocated mainly by these authors. This model is reasonable, but it is widely disputed (e.g., Glanzer, Kim, Hilford, & Adams, 1999; Heathcote, 2003). Thus, it should be understood that the validity of any conclusions reached by these further analyses are dependent on the validity of a model that remains controversial. For example, this model was used to interpret the recall and recognition  $z$  scores by subjecting them to a *path analysis* and was used to interpret remember/know and receiver operating characteristic (ROC) data from two additional and smaller experiments as well.

### Remember/Know and ROC Analyses

Both Yonelinas et al. (2002) and Manns et al. (2003) reported a second experiment in which they attempted to measure the familiarity and recollection processes in hippocampal patients, using the remember/know procedure. In this procedure, subjects complete a standard recognition test in which they decide whether test items are *old* or *new*, and for each *old* recognition decision, they further indicate whether their response was based on recollection (remember) or familiarity (know). Here, the two studies also came to different conclusions.

An interpretation of remember/know judgments is not possible without reference to a specific theory of recognition memory, and what the appropriate theory might be remains controversial. Some have suggested that these judg-

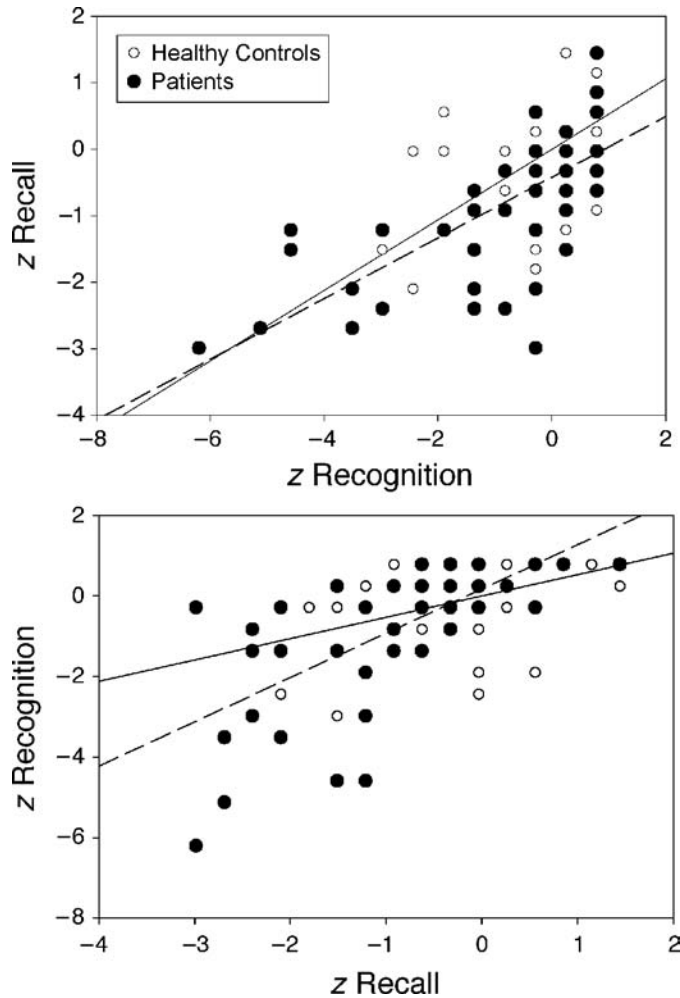


Figure 4.  $z$  recall versus  $z$  recognition scores (upper panel) and  $z$  recognition versus  $z$  recall scores (lower panel) for patients (filled symbols) and healthy controls (open symbols) (data from Yonelinas et al., 2002, with the outlier excluded). Some symbols represent data for more than 1 individual who received the same recall and recognition scores. The solid line represents the regression line for the controls, and the dashed line represents the regression line for the patients.

ments do not measure different underlying processes (e.g., Donaldson, 1996; Dunn, 2004; Inoue & Bellezza, 1998) and, instead, merely reflect different strengths of memory. This interpretation, which is based on the standard signal detection view of recognition memory, holds that when the memory strength of a test item exceeds a high criterion, a remember response is given. If the strength of a test item exceeds only a lower criterion, a know response is given.

Others have suggested that remember/know judgments are more than just indicators of memory strength and that they do provide valid measures of recollection and familiarity (Gardiner & Java, 1990; Yonelinas et al., 1998). Yonelinas and his colleagues have proposed a quantitative dual-process model of recognition memory that provides one way to interpret remember/know judgments in order to arrive

at estimates of recollection and familiarity (e.g., Yonelinas, 2002). The recollection estimate is based on the standard high-threshold *correction for guessing* formula and is equal to  $(R_{HR} - R_{FA}) / (1 - R_{FA})$ , where  $R_{HR}$  and  $R_{FA}$  are the remember hit and false alarm rates, respectively. The familiarity estimate is a  $d'$  value computed from the hit and false alarm rates for know judgments, after first adjusting those rates to take into consideration the proportion of remember responses. Specifically, the  $d'$  estimate is based on adjusted know hit rates and false alarm rates derived from the following equations:

$$K'_{HR} = K_{HR} / (1 - R_{HR})$$

and

$$K'_{FA} = K_{FA} / (1 - R_{FA}),$$

where  $K'_{HR}$  and  $K'_{FA}$  represent the adjusted know hit rates and false alarm rates, respectively, and  $K_{HR}$  and  $K_{FA}$  represent the unadjusted know hit rates and false alarm rates, respectively. The dependent measure that is computed from  $K'_{HR}$  and  $K'_{FA}$  is  $d'$ , because the familiarity process that gives rise to know responses is assumed to be accurately characterized by a signal detection model.

Because the familiarity measure is a  $d'$  score and the recollection measure is a probability, the two scales are not directly comparable. Yonelinas et al. (2002) avoided this problem by not computing a  $d'$  score (as their theory requires) and using, instead, a contrived familiarity measure consisting of the corrected know hit rate minus the corrected know false alarm rate (i.e.,  $K'_{HR} - K'_{FA}$ ). They also used an abbreviated equation to estimate recollection (which consisted of  $R_{HR} - R_{FA}$ ), instead of using the full high-threshold correction formula. When these measures of familiarity and recollection are used, the data reported by Yonelinas et al. (2002) appear to suggest a differential deficit, with recollection more impaired than familiarity. The recollection measures for the controls and the patients were .39 and .06, respectively, whereas the familiarity measures for the controls and the patients were .32 and .28, respectively.

Table 1 presents the average remember and know hit and false alarm rates for the patients and the controls in both studies. The values in the left two columns were made available to us by the authors of Yonelinas et al. (2002), and the values in the right two columns were taken from Manns et al. (2003). The table also presents a quantitative estimate of the processes that are assumed to underlie remember (recollection) and know (familiarity) responses according to the model advocated by Yonelinas and his colleagues. The recollection estimate was computed using the high-threshold formula, and the familiarity estimate is a  $d'$  value computed from the corrected know hit and false alarm rates ( $K'_{HR}$  and  $K'_{FA}$ , respectively). These estimates differ from the ones reported by Yonelinas et al. (2002), because the authors used abbreviated estimates instead of the quantitative estimates suggested by their model.

These process estimates suggest that the disagreement between the two studies may be more apparent than real.

The values in Table 1 show that the  $d'$  (familiarity) score is more than 23% lower for the patients relative to the controls in Yonelinas et al. (2002), which weighs against the idea that familiarity-based responding is unimpaired in these patients. The effect may not be statistically significant, but with only 4 patients that is not surprising. Thus, the main difference between the data reported by Yonelinas et al. (2002) and the data reported by Manns et al. (2003) is in the apparent size of the familiarity deficit (23% lower score vs. 69% lower score). This difference may be the result of the large measurement error that is associated with small sample sizes. With larger sample sizes, the two studies might have revealed more similar familiarity deficits.

Whether or not this is the case, other aspects of the remember/know data would appear to prohibit drawing any strong conclusions. If hippocampal patients have severely impaired recollection and if remember responses reflect recollection, one would expect to see a paucity of such responses to target items and none at all to lure items. But the remember hit and false alarm rates reported by Yonelinas et al. (2002) were .23 and .16, respectively. That is, even for the lures, the patients claimed to have remembered 16% of them. The remember false alarm rate for Manns et al.'s (2003) hippocampal patients was also .16. These high remember false alarm rates complicate the interpretation of remember/know data.

Yonelinas et al. (2002) also reported a third experiment in which they examined the ROC of 3 of the 4 patients who participated in the remember/know experiment. An ROC is a plot of the hit rate versus the false alarm rate for different levels of confidence associated with recognition decisions. Traditionally, ROC data are analyzed on the basis of signal detection theory (e.g., Ratcliff, Sheu, & Gronlund, 1992), but Yonelinas et al. (2002) analyzed their ROC data according to the same dual-process model that they had used to interpret remember/know judgments. According to this model, most high-confident recognition decisions reflect responses based on an all-or-none recollection process, whereas all decisions associated with low and medium levels of confidence reflect responses based on varying degrees of familiarity. Whereas decisions based on familiarity are assumed to be accurately characterized by a signal detection model, recollection is thought to arise from an altogether different threshold process that allows for no degrees of recollective strength (i.e., recollection is all or none). When the ROC data were analyzed in terms of that model, the results again suggested that, relative to controls, recollection was severely impaired in the patients but the familiarity process was intact.

This conclusion, which is based on very few patients, hinges on the validity of the quantitative dual-process model that the authors fit to their data. Whereas that model assumes that recognition decisions are based either on recollection or on familiarity (never both), a more traditional signal detection model assumes that recognition decisions are based on the strength of the memory signal relative to a criterion (Lockhart & Murdock, 1970; Ratcliff et al.,

**Table 1**  
**Hit Rates, False Alarm Rates, and Memory Process Estimates for Remember (R) and Know (K) Judgments in Yonelinas et al. (2002) and Manns, Hopkins, Reed, Kitchener, and Squire (2003)**

Measure	Yonelinas et al.		Manns et al.	
	R	K	R	K
Hippocampal Patients				
Hit rate	.23	.49	.35	.22
False alarm rate	.16	.30	.16	.19
Process estimate	.08	0.72	.23	0.33
Healthy Controls				
Hit rate	.43	.27	.44	.25
False alarm rate	.04	.15	.05	.11
Process estimate	.41	0.94	.41	1.06

1992). The memory signal generated by a test item may arise from a single process (e.g., familiarity), or it may be a joint function of multiple processes (e.g., recollection and familiarity), but responding is always assumed to be based on the strength of that signal no matter how many processes contribute to it. The standard signal detection model holds that the memory strengths associated with the targets and lures are normally distributed, with the mean of the target distribution being situated higher on the strength axis than the mean of the lure distribution. An *old/new* decision criterion is also placed somewhere along the strength axis, and any item that generates a memory signal that exceeds the decision criterion is declared to be *old* (otherwise it is declared to be *new*).

The question of whether signal detection theory or the dual-process model better fits ROC data has been (and continues to be) strongly debated (Glanzer, Hilford, Kim, & Adams, 1999; Glanzer et al., 1999; Yonelinas, 1999). Most recently, Heathcote (2003) presented evidence that, for standard item recognition tests, the detection model almost always outperforms the dual-process model advocated by Yonelinas and his colleagues. That being the case, it seems reasonable to ask what the ROC data reported by Yonelinas et al. (2002) would mean if they were interpreted in terms of signal detection theory.

As is shown in Figure 4b of Yonelinas et al. (2002), the ROC data produced by the healthy controls trace out a curvilinear path well above the diagonal, whereas the ROC data produced by the patients trace out a path much closer to the diagonal (with the diagonal representing chance performance). According to signal detection theory, findings such as these indicate different levels of memory strength. That is, an ROC curve that falls close to the diagonal corresponds to a detection model in which the mean of the target distribution is close to the mean of the lure distribution (i.e., on average, the targets and lures do not differ much in terms of strength). An ROC curve farther above the diagonal corresponds to a detection model in which the mean of the target distribution is substantially higher than the mean of the lure distribution. From this point of view, the ROC results reported by Yonelinas et al. (2002) simply confirm what Manns and Squire (1999) previously reported—namely, that recognition memory is impaired in patients with damage limited to the hippocampus. This is important because some have argued otherwise (Mayes, Holdstock, Isaac, Hunkin, & Roberts, 2002). If recognition memory had been spared in the patients tested by Yonelinas et al. (2002), the ROC data for the hippocampal patients and healthy controls should have been indistinguishable. Instead, a standard detection analysis of the ROC data suggests a marked impairment in the recognition performance of the hypoxic patients. The detection analysis is silent on the question of whether the impairment is selective to one subprocess of declarative memory.

### Conclusion

The data reported by Manns et al. (2003) and Yonelinas et al. (2002) appear to be in disagreement as to whether or

not recall is more impaired than recognition in patients with damage restricted to the hippocampus. However, a close inspection of the individual data reveals that the disagreement arises mainly because 1 of the 55 controls in Yonelinas et al.'s (2002) study had an aberrantly low recognition score. When that single outlying score is removed, the results of both studies suggest that recall and recognition are equivalently impaired in patients with hippocampal lesions. This finding is important because a comparison of recall and recognition is probably the most direct way to test whether or not different subprocesses of declarative memory are differentially impaired in patients with amnesia. The test is compelling because it relies only on the common assumption that recognition is based on recollection and familiarity, whereas recall is based only on recollection. Other methods, such as the remember/know procedure and ROC analysis, rely on specific quantitative models that involve assumptions that are not widely shared. Further complicating the remember/know procedure is the fact that the instructions need to be quite detailed, and concern is often expressed that subjects fail to understand them (which may be one reason why amnesic patients often produce high remember false alarm rates).

The conclusions drawn by Yonelinas et al. (2002) are further weakened by the fact that no anatomical information is available for any of the 56 hypoxic patients they tested in their recall/recognition experiment or for the 3 to 4 patients they tested in their remember/know and ROC analyses. As was indicated above, the authors supposed that a period of hypoxia results in damage limited to the hippocampal region. However, the presumption of selective hippocampal damage in the absence of radiologic or histological data is unwarranted. Thus, in one study (Grubb et al., 2000), an MRI analysis of the brains of cardiac arrest survivors with evidence of memory impairment revealed global cerebral atrophy, rather than selective hippocampal damage. Whether or not the same is true of the patients tested by Yonelinas et al. (2002) is not known. Note, however, that Manns et al. (2003) reported quantitative data from MRI scans for their patients. These scans showed volume reductions for the hippocampal region of approximately 40%, on average, as well as a negligible average reduction (3%) in the volume of the parahippocampal gyrus.

On balance, the findings reported by Yonelinas et al. (2002) do not suggest that damage limited to the hippocampal region selectively impairs recollection while leaving familiarity largely intact. Instead, the weight of evidence suggests that damage to the hippocampal region impairs declarative memory broadly—recall and recognition and both recollection and familiarity.

### REFERENCES

- BROWN, M., & AGGLETON, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*, 51-61.
- DONALDSON, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*, 523-533.



- DUNN, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, **111**, 524-542.
- GARDINER, J. M., & JAVA, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, **18**, 23-30.
- GLANZER, M., HILFORD, A., KIM, K., & ADAMS, J. K. (1999). Further tests of dual-process theory: A reply to Yonelinas. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 522-523.
- 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, & Cognition*, **25**, 500-513.
- GRUBB, N. R., FOX, K. A., SMITH, K., BEST J., BLANE, A., EBMEIER, K. P., GLABUS, M. F., & O'CARROLL, R. E. (2000). Memory impairment in out-of-hospital cardiac arrest survivors is associated with global reduction in brain volume, not focal hippocampal injury. *Stroke*, **31**, 1509-1514.
- HAIST, F., SHIMAMURA, A. P., & SQUIRE, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 691-702.
- HEATHCOTE, A. (2003). Item recognition memory and the ROC. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1210-1230.
- INOUE, C., & BELLEZZA, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition*, **26**, 299-308.
- LOCKHART, R. S., & MURDOCK, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, **74**, 100-109.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.
- MANN, J. R., HOPKINS, R. O., REED, J. R., KITCHENER, E. G., & SQUIRE, L. R. (2003). Recognition memory and the human hippocampus. *Neuron*, **38**, 127-133.
- MANN, J. R., & SQUIRE, L. R. (1999). Impaired recognition memory on the doors and people test after damage limited to the hippocampal region. *Hippocampus*, **9**, 495-499.
- MAYES, A. R., HOLDSTOCK, J. S., ISAAC, C. L., HUNKIN, N. M., & ROBERTS, N. (2002). Relative sparing of item recognition memory in a patient with adult-onset damage limited to the hippocampus. *Hippocampus*, **12**, 325-340.
- RATCLIFF, R., SHEU, C. F., & GRONLUND, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, **99**, 518-535.
- RUGG, M. D., & YONELINAS, A. P. (2003). Human recognition memory: A cognitive neuroscience perspective. *Trends in Cognitive Sciences*, **7**, 313-319.
- SCHACTER, D. L., & TULVING, E. (1994). *Memory systems*. Cambridge, MA: MIT Press.
- SQUIRE, L. R. (1992). Memory and the hippocampus: A synthesis of findings with rats, monkeys and humans. *Psychological Review*, **99**, 195-231.
- TUKEY, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- YONELINAS, A. P. (1999). Recognition memory ROCs and the dual-process signal-detection model: Comment on Glanzer, Kim, Hilford, and Adams. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 514-521.
- YONELINAS, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language*, **46**, 441-517.
- YONELINAS, A. P., KROLL, N. E. A., DOBBINS, I., LAZZARA, M. M., & KNIGHT, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, **12**, 323-339.
- YONELINAS, A. P., KROLL, N. E. A., QUAMME, J. R., LAZZARA, M. M., SAUVE, M., WIDAMAN, K. F., & KNIGHT, R. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience*, **5**, 1236-1241.

#### NOTE

1. The recognition scores for Trials 2-5 were at the ceiling for the controls, so that a meaningful  $z$  score analysis for those trials was not possible.

(Manuscript received October 1, 2003;  
revision accepted for publication December 14, 2003.)