

Mild hypoxia disrupts recollection, not familiarity

A. P. YONELINAS, J. R. QUAMME, K. F. WIDAMAN, and N. E. A. KROLL
University of California, Davis, California

M. J. SAUVÉ
University of California Davis Medical Center, Sacramento, California

and

R. T. KNIGHT
University of California, Berkeley, California
and Helen Wills Neuroscience Institute, Berkeley, California

Yonelinas et al. (2002) found that hypoxic patients exhibited deficits in recollection that left familiarity relatively unaffected. In contrast, Manns, Hopkins, Reed, Kitchener, and Squire (2003) studied a group of hypoxic patients who suffered severe and equivalent deficits in recollection and familiarity. We reexamine those studies and argue that the discrepancy in results is likely due to differences in the hypoxic groups that were tested (i.e., differences in amnesic severity, subject sampling methods, and patient etiology). Yonelinas et al. examined memory in 56 cardiac arrest patients who suffered a brief hypoxic event, whereas Manns et al. examined a group of severely amnesic patients that consisted of 2 cardiac arrest patients, 2 heroin overdose patients, 1 carbon monoxide poisoning patient, and 2 patients with unknown etiologies. We also consider an alternative explanation proposed by Wixted and Squire (2004), who argued that the two patient groups suffered similar deficits, but that statistical or methodological artifacts distorted the results of each of Yonelinas et al.'s experiments. A consideration of those results, however, indicates that such an explanation does not account for the existing data. All of the existing evidence indicates that recollection, but not familiarity, is disrupted in mild hypoxic patients. In more severe cases of hypoxia, or those with more complex etiologies such as heroin overdose, more profound deficits may be observed.

Recognition memory judgments can be based on recollection of qualitative information about a previous study event or on assessments of stimulus familiarity. Two competing classes of theories have been proposed to account for the contribution of the medial temporal lobes to recollection and familiarity. According to one *dual-process* account, the hippocampus is involved in recollection, whereas the parahippocampal gyrus is involved in familiarity (e.g., Aggleton & Brown, 1999; Eichenbaum, Otto, & Cohen, 1994). By another account, both recollection and familiarity reflect forms of *declarative memory* and are equally dependent on a single medial temporal lobe system (e.g., Squire, 1994; Squire & Zola, 1998). Critical in testing these opposing views are recent results from studies of hypoxic patients, because imaging and histological research have suggested that damage in these patients can have occurred primarily in the hippocampus (e.g., Gadian et al., 2000; Rempel-Clower, Zola, Squire, & Amaral, 1996; but see Caine & Watson, 2000).

Results from two recent studies have led to opposing conclusions regarding the fate of recollection and familiarity in hypoxia. Yonelinas et al. (2002) found that esti-

mates of recollection, but not familiarity, were decreased by hypoxia. In contrast, Manns, Hopkins, Reed, Kitchener, and Squire (2003) found that estimates of recollection and familiarity were reduced severely, and to an equal extent, by hypoxia. In the present paper, we examine those two studies and argue that the conflicting results likely reflect important differences in the patient groups examined in those studies. We also consider an artifact explanation (cf. Wixted & Squire, 2004) and show that it does not provide an adequate account of the results.

Methods of Estimating Recollection and Familiarity

To assess the effects of hypoxia on recollection and familiarity, it is necessary to estimate the contribution of these two processes to memory performance (for a review of methods, see Yonelinas, 2002). Early attempts to assess these two processes relied on *task dissociation methods*, in which one tried to find a task that provided an index of recollection and another that provided an index of familiarity. For example, if one assumes that performance on a free recall test provides a pure measure of recollection and that performance on a recognition test provides a measure of familiarity, then patients with selective recollection deficits should have greater deficits

Correspondence concerning this article should be addressed to A. P. Yonelinas, Department of Psychology, University of California, Davis, CA 95616 (e-mail: apyonelinas@ucdavis.edu).

in recall than in recognition. However, it is now generally agreed that memory tasks do not provide direct measures of underlying memory processes (see, e.g., Jacoby, Toth, & Yonelinas, 1993; Moscovitch, 1994; Roediger, 1990; Tulving, 1985). For example, recognition can be based on assessments of either recollection or familiarity, and the relative contributions of these two processes can vary quite widely from experiment to experiment. Moreover, although recollection is likely to be a major contributor to recall tests, familiarity may also lead items to come to mind in such a test, thus further clouding the interpretation of these task comparisons.

Dissatisfaction with task dissociation methods has led to the development of methods to directly measure the memory processes themselves (i.e., *process dissociation* methods). The aim is to derive parameter estimates of the underlying memory processes that give rise to memory performance, rather than to equate different memory tasks directly with distinct memory processes. For example, using the remember/know procedure (e.g., Gardiner, 1988; Tulving, 1985; Yonelinas & Jacoby, 1995), subjects are asked to indicate which of their recognition responses are based on recollection of qualitative information about the study event and which are based on familiarity in the absence of recollection. Recollection is then estimated by examining the proportion of remember responses, whereas familiarity is estimated as the proportion of familiarity-based responses in the absence of recollection. Another method is to fit a quantitative model to confidence-based receiver operating characteristics (ROCs; Yonelinas, 1994, 2001). Subjects are required to rate their confidence in their recognition responses, and hits (i.e., proportion of studied items correctly recognized) are plotted against false alarms (i.e., the proportion of nonstudied items incorrectly recognized) as a function of response confidence. Fitting the observed function to a quantitative model of performance can provide estimates of recollection and familiarity. The method is analogous to conducting a linear regression to derive slope and intercept estimates. Another approach is to use structural equation modeling to examine the underlying processes (i.e., latent variables) that give rise to overt performance (e.g., Jöreskog, 1974; Nyberg, 1994). With this method, one can examine the effects of hypoxic severity on latent variables, such as recollection and familiarity, that underlie recall and recognition performance.

Process estimation methods have been used quite extensively over the past 20 years to assess recollection and familiarity, and this has led to a remarkably consistent picture of the functional characteristics and neural substrates of these two forms of memory (for reviews, see Rugg & Yonelinas, 2003; Yonelinas, 2002). However, an important limitation of these estimation methods is that they rely on various critical assumptions concerning how the underlying processes are related, such as, that subjects can reliably report on these processes, that recollection leads to high-confidence recognition responses, or that latent factors contribute linearly to observed performance. Such assumptions can of course be questioned,

and one must be cautious in interpreting results from any of these methods. However, these limitations can be overcome by using test conditions under which the assumptions are likely to be met, conducting analyses in which the critical assumptions are directly tested, and relying on convergent operations whereby the results of one method are confirmed using a variety of alternative methods.

Estimates of Recollection and Familiarity in Hypoxia

In two recent studies, researchers examining the effects of hypoxia on recollection and familiarity have arrived at quite different conclusions. In one of them, using three different measurement methods, Yonelinas et al. (2002) found that hypoxic patients exhibited deficits in recollection, with familiarity relatively unaffected. First, structural equation modeling of recall and recognition performance in 56 mildly hypoxic cardiac arrest patients indicated that increased hypoxic severity led to a reduction in estimates of recollection, but did not influence familiarity. Second, remember/know reports in a smaller recognition experiment provided convergent results showing that the hypoxic patients suffered a deficit in recollection, but not familiarity (also see Kishiyama, Yonelinas, & Lazzara, 2004, for similar results with a different set of materials). Third, ROCs obtained in another experiment also showed that the patients suffered a selective deficit in recollection.

In addition to using these process estimation methods, Yonelinas et al. (2002) compared the recall and recognition deficits of the patients with those of a group of healthy control subjects. Relative to the controls, the patients exhibited reduced recall and recognition scores, but their recall deficit was marginally greater than their recognition deficit. The finding that recall was more impaired than recognition is inconsistent with single-system declarative models of memory, in which both recognition and recall are treated as measures of the same underlying memory system and thus should be equally impaired (e.g., Squire, 1994). However, as Yonelinas et al. pointed out, the result is not particularly relevant in testing the dual-process account, because it does not provide estimates of recollection and familiarity. Thus, it cannot tell us whether the patients suffered selective deficits in recollection. Moreover, even a failure to find a significantly greater recall deficit than recognition deficit would not be particularly informative for the model. Such a finding could simply be due to the fact that both the recall and recognition tests relied heavily on recollection. Unless estimates of recollection and familiarity are derived for both recall and recognition memory tests, it is not possible to make any clear predictions about the relative impairments seen in these tests. Indeed, given the likelihood that recollection and familiarity will have differential contributions to performance based on the particular memory paradigm, any comparison between the manifest variables of recall and recognition will be a shifting target, whereas direct tests of differential effects

on the underlying process variables of recollection and familiarity will not suffer from this confound.

In a separate study, Manns et al. (2003) used the remember/know procedure to measure recollection and familiarity in 7 hypoxic patients and found that hypoxia led to comparable, and quite severe, impairments in estimates of recollection and familiarity. They also found comparable deficits in recall and recognition performance, which they interpreted as indicating that both recollection and familiarity were equally impaired. Although the latter finding is not particularly useful in determining whether recollection is selectively impaired in hypoxic patients because no measures of recollection and familiarity were derived in those tests, the remember/know results are relevant, and they do appear to differ from those of Yonelinas et al. (2002). Why did the results of the two studies differ?

Differences Between Patient Groups

The patient groups in the preceding two studies differed in at least three important ways (i.e., amnesic severity, the manner in which subjects were selected, and etiology), and any one of these differences could have been responsible for the differing conclusions of those two studies. First, the patients in Manns et al. (2003) were much more severely memory impaired than the patients in Yonelinas et al. (2002). For example, the delayed recall scores on the RAVLT in Yonelinas et al.'s study were, on average, less than one *SD* below those of the controls (recall $z = -0.68$; recognition $z = -0.39$), and the worst recall score in the group of 56 patients was -2.99 . Manns et al. did not report delayed recall scores on the RAVLT; but they did report performance in an immediate condition in the RAVLT in which there was no delay between study and test, and their patients were already close to 2 *SDs* below normal (recall $z = -1.83$; recognition $z = -1.91$). In delayed tests of long-term memory, the Manns et al. patients performed so poorly that they were essentially off the scale. For example, 6 of the 7 patients produced scores of <50 on the delayed WMS-R (Wechsler Memory Scale-Revised) memory measures. On this test, 100 is the population mean, with an *SD* of 15 in the population. Note that the norming tables for this test do not go lower than 50, so these scores likely underestimate the severity of the delayed memory impairments in the Manns et al. patients. Moreover, on another standardized test (the Doors and People Test) the patients also performed exceedingly poorly. Although the normed scores were not provided for each of the patients, the scores were presented for 3 of the patients in an earlier study (Manns & Squire, 1999), where the scores were lower than the 1st percentile. Again, the raw scores of these patients fell below the range for which the testing manual provided normative data, and thus these scores likely underestimate their impairments. As described below, there is good reason to suspect that more severe hypoxic patients exhibit more pronounced pathology, which may well lead to deficits in familiarity.

Second, there appear to be critical differences in the way the patient groups were selected. The patients in Yonelinas et al. (2002) were not selected on the basis of their memory impairments, but rather on the basis of their etiology: cardiac arrest. This is a critical point because it allowed Yonelinas et al. to examine the effects of cardiac arrest on subsequent memory. In contrast, the patients in Manns et al. (2003) appear to have been selected to include only those with extremely low memory scores. This would explain why there were no patients with mild memory impairments, and it might explain why both recollection and familiarity were impaired in their patient group. For example, even in a group of normal subjects, if one selected the subjects with the poorest memory scores, it might not be surprising to find that they also exhibited low recollection and familiarity scores.

Third, the studies also differed with respect to the patient etiologies. All the patients in Yonelinas et al. (2002) were cardiac arrest patients, reflecting homogeneity with regard to etiology of the memory impairment. Presumably the brain damage and memory impairments were directly related to the temporary reduction in blood flow due to the cardiac arrest. In contrast, the patient group in Manns et al. (2003) consisted of 2 cardiac arrest patients, 2 heroin overdose patients, 1 carbon monoxide poisoning patient, and 2 patients with unknown etiologies. Both heroin overdose and CO₂ poisoning produce neurotoxic effects beyond those typically related to hypoxia (e.g., Kono, Kono, & Shida, 1983; O'Donnell, Buxton, Pitkin, & Jarvis, 2000; Pearson, Baden, & Richter, 1976; Yee, Gronner, & Knight, 1994). Although the recall and recognition deficits of the different types of patients do not appear to be very different, with only 1 or 2 patients from each subgroup such a comparison cannot be considered conclusive.

The neuropathological basis of the familiarity deficits seen in the Manns et al. (2003) patients is not yet clear. However there are several possibilities. For example, familiarity may be impaired in their patients relative to mild hypoxics because they suffered more complete damage to the hippocampus, or because they suffered damage to different regions within the hippocampus. Alternatively, the familiarity deficits could be related to damage to regions outside the hippocampus. Consistent with this is the observation that as hypoxic severity increases, the likelihood of damage outside the hippocampus increases (e.g., Kono et al., 1983; Smith, Auer, & Siesjo, 1984). Manns et al. reported extensive hippocampal atrophy in most patients, and only minor (15% in the worst case) reductions in the surrounding parahippocampal cortex, but no other volumetric results were reported. Familiarity-based recognition has been related to a variety of regions throughout the brain such as the thalamus, the prefrontal cortex, and the parietal cortex (e.g., Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000; Henson, Rugg, Shallice, Josephs, & Dolan, 1999; Kishiyama et al., 2004; Ranganath et al., 2004; Wheeler & Stuss, 2003; Zoppelt, Koch, Schwarz, & Daum, 2003). To determine whether damage to these other brain

regions is responsible for the observed familiarity deficits in Manns et al., it will be important to determine whether their patients suffered structural damage or functional disruption to these brain regions.

Whether the additional deficits seen in Manns et al.'s (2003) patients are related to amnesic severity, the method used to select those subjects, or the various etiologies that were included, is difficult to determine without further study. Nevertheless, these differences in patient groups cannot be overlooked when one interprets the differences between the two studies, and, as just described, there is reason to suspect that the differences played a critical role.

Alternative Accounts

An alternative account of the discrepancy between the two studies was recently proposed by Wixted and Squire (2004), who argue that both patient groups exhibited severe impairments in recollection and familiarity, but that Yonelinas et al.'s (2002) experiments were distorted by a variety of different artifacts that made it look as if their patients were selectively impaired at recollection. These "artifact" arguments are considered below and are found to be insufficient to account for the observed patterns of results.

Remember/know results. With respect to the remember/know results, Wixted and Squire (2004) suggested that the hypoxic patients in both studies may have misunderstood the remember/know instructions. They pointed out that the levels of "remember" false alarms were quite high for the hypoxics in the Yonelinas et al. (2002) and Manns et al. (2003) remember/know experiments, suggesting that the patients may have been making remember responses even when they did not truly remember the items. However, the Yonelinas et al. research group went to considerable lengths to ensure that the subjects understood the remember/know distinction by having them repeat the test instructions, and by requiring them to provide justification for their remember and know responses at points throughout the test phase. Although we cannot speak for Manns et al., there was no indication that the hypoxic patients misunderstood the instructions in Yonelinas et al. More importantly, if the patients sometimes made a remember response when an item was only familiar, this would have reduced the familiarity estimates of the patients. This might explain why the hypoxics in Manns et al. had low familiarity estimates, but it does not explain why the patients in Yonelinas et al. showed normal familiarity estimates. Thus, the "misunderstanding" artifact could not have produced the discrepancy between the two studies.

Structural modeling. Wixted and Squire (2004) further suggested that the recall and recognition results in Yonelinas et al. (2002) may have been biased by the inclusion of a single outlier in the control group. However, the inclusion or exclusion of this subject is irrelevant, because the estimates of recollection and familiarity that we derived were based solely on the performance of the patients. The control data were never used in the struc-

tural modeling, so the single "outlying" control subject could not have affected those estimates.

Including the outlying control subject can have an effect when one contrasts the recall and recognition scores of the patients and controls. Specifically, the extent to which recall is more impaired than recognition is somewhat increased when the subject is included. However, regardless of the outlier, a majority of the hypoxic patients exhibited a more pronounced recall deficit than recognition deficit, confirming previous studies of amnesia (e.g., Hirst, Johnson, Phelps & Volpe, 1988; Mayes, Holdstock, Isaac, Hunkin, & Roberts, 2002; Vargha-Khadem et al., 1997; but see Haist, Shimamura, & Squire, 1992; Manns et al., 2003). In any case, as has been described above, this task comparison tells us little about recollection and familiarity.

A potentially more important point made by Wixted and Squire (2004) is that the outlier and high levels of performance (i.e., ceiling effects) on the recall and recognition tests may have biased the estimates derived in the structural modeling. It is important to point out that, contrary to Wixted and Squire's assumption, control data were never used to derive estimates of recollection and familiarity in the structural equation modeling, and ceiling effects were much less of a problem in the patients (i.e., only in recognition was there evidence of ceiling effects). Nevertheless, we took their suggestion quite seriously and asked whether ceiling effects in the patients could have affected the structural modeling results in such a way that it led us to mistaken conclusions.

To explore this possibility, we simulated data from a single-process memory model (cf. Squire, 1994) in which recall and recognition reflected a single underlying declarative memory system and in which recognition scores suffered from ceiling effects in the same way as was observed in the Yonelinas et al. (2002) data (see the Appendix for details of the analysis). We then looked to see whether the ceiling effects could have artifactually created a second memory component (i.e., familiarity) that was not affected by hypoxia. Our simulation results showed that, although ceiling effects might worsen slightly the overall fit of a single-component model, they do not cause false, extra factors to be identified. Thus, these simulation results argue strongly against the position that the structural equation modeling results were distorted by ceiling effects.

Wixted and Squire (2004) also suggested that structural equation modeling results could not be trusted because they are based on the validity of the assumptions underlying a particular model. It is true that all measurement methods make assumptions. However, one advantage of the structural equation method is that one can directly test the assumptions underlying the model and can directly contrast a preferred model with alternative models. Quamme, Yonelinas, Widaman, Kroll, and Sauvé (2004) recently examined the fit of the dual-process model to the hypoxic data and found that the model provided an excellent account. Moreover, direct tests indi-

cated that the model's core assumptions were correct. Specifically, there were two independent factors underlying recognition, and only one of these (i.e., recollection) was affected by coma duration. Finally, direct contrasts indicated that the dual-process model provided a better fit to the data than did a variety of alternative models, including a single-component declarative memory model (cf. Squire & Zola, 1998).

ROC estimates. Finally, Wixted and Squire (2004) argued that the results of the ROC experiment could be questioned because this analysis relied on the assumptions of a particular ROC model. Evidence demonstrating the validity of the ROC model has been discussed at length previously (e.g., Yonelinas, 1999, 2001, 2002) and will not be reiterated here. We should, however, point out that the two references provided by Wixted and Squire as containing evidence against that ROC model actually showed that, along the entire ROC, the predicted hit or false alarm rate deviated from the observed scores by no more than 1 or 2 percent (e.g., Glanzer, Kim, Hilford, & Adams, 1999; Heathcote, 2003). Interestingly, the deviation was almost always in the direction that one would expect if subjects occasionally made random responses (Ratcliff, McKoon, & Tindall, 1994; Yonelinas, 1999)—a very real possibility in ROC experiments, in which subjects are required to make hundreds of recognition responses. In any case, such a minor deviation can hardly be considered a major challenge to the ROC analysis.

The importance of convergence. The aspect of the Yonelinas et al. (2002) study that we find most compelling in arguing against the multiple artifact arguments of Wixted and Squire (2004) is that the conclusions were supported by convergent findings from several very different analytic methods. Regardless of whether one operationalized recollection and familiarity as independent sources of variance underlying recognition, as introspective reports of retrieval experience, or as ROC curve parameters in a mathematical model of response confidence, the experiments showed that mild hypoxia reduced recollection, but failed to influence familiarity. One can always criticize the assumptions of a given method or experiment and attribute the successful prediction of any particular effect to artifacts to which the method is susceptible. However, when estimates derived from very different methodological and analytical procedures converge consistently in favor of a particular theory, it becomes increasingly unlikely that the success of the theory in accounting for observations is the result of an artifact. Indeed, at some point, it must be conceded that the particular collection of artifacts necessary to explain the range of data is so improbable that one has the choice of either acknowledging the theory's merit, or "attributing replicable orderliness of observations to a damn strange coincidence" (Meehl, 1990, p. 117). The claim that mild hypoxia is related to a deficit in recollection that leaves familiarity relatively unaffected is supported by a wide variety of results. We suggest that this convergence is more than a "damn strange coincidence."

Conclusions

The discrepancy between the results of the Yonelinas et al. (2002) and Manns et al. (2003) studies is an important one that is not easily explained away as arising from measurement artifacts. We have argued that it likely is related to differences in the patient groups that were examined in the two studies. The preponderance of evidence indicates that, in mild cases of hypoxia induced by cardiac arrest, recollection, but not familiarity, is impaired. In contrast, in more severely amnesic patients or in patients having more complex etiologies, such as heroin overdose or carbon monoxide poisoning, it appears that both recollection and familiarity can be impaired.

REFERENCES

- AGGLETON, J. P., & BROWN, M. W. (1999). Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. *Behavioral & Brain Sciences*, *22*, 425-444.
- CAINE, D., & WATSON, J. D. G. (2000). Neuropsychological and neuropathological sequelae of cerebral anoxia: A critical review. *Journal of the International Neuropsychological Society*, *6*, 86-99.
- EICHENBAUM, H., OTTO, T., & COHEN, N. J. (1994). Two functional components of the hippocampal memory system. *Behavioral & Brain Sciences*, *17*, 449-517.
- ELDRIDGE, L. L., KNOWLTON, B. J., FURMAN, C. S., BOOKHEIMER, S. Y., & ENGEL, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, *3*, 1149-1152.
- GADIAN, D. G., AICARDI, J., WATKINS, K. E., PORTER, D. A., MISHKIN, M., & VARGHA-KHADEM, F. (2000). Developmental amnesia associated with early hypoxic-ischaemic injury. *Brain*, *123*, 499-507.
- GARDINER, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309-313.
- 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.
- 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 receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 1210-1230.
- HENSON, R. N. A., RUGG, M. D., SHALLICE, T., JOSEPHS, O., & DOLAN, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, *19*, 3962-3972.
- HIRST, W., JOHNSON, M. K., PHELPS, E. A., & VOLPE, B. T. (1988). More on recognition and recall in amnesics. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 758-762.
- ISAAC, C. L., & MAYES, A. R. (1999). Rate of forgetting in amnesia: I. Recall and recognition of prose. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 942-962.
- JACOBY, L. L., TOTH, J. P., & YONELINAS, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*, 139-154.
- JÖRESKOG, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Measurement, psychophysics, and neural information processing* (pp. 1-56). San Francisco: Freeman.
- KISHIYAMA, M. M., YONELINAS, A. P., & LAZZARA, M. M. (2004). The von Restorff effect in amnesia: The contribution of the hippocampal system to novelty-related memory enhancements. *Journal of Cognitive Neuroscience*, *16*, 15-23.
- KONO, E., KONO, R., & SHIDA, K. (1983). Computerized tomographies of 34 patients at the chronic stage of acute carbon monoxide poisoning. *Archiv für Psychiatrie und Nervenkrankheiten*, *233*, 271-278.
- MANN, J. R., HOPKINS, R. O., REED, J. M., KITCHENER, E. G., & SQUIRE,

- L. R. (2003). Recognition memory and the human hippocampus. *Neuron*, **27**, 171-180.
- MANNES, 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.
- MEEHL, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, **1**, 108-141.
- MOSCOVITCH, M. (1994). Memory and working with memory: Evaluation of a component process model and comparisons with other models. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 267-310). Cambridge, MA: MIT Press.
- NYBERG, L. (1994). A structural equation modeling approach to the multiple memory systems question. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 485-491.
- O'DONNELL, P., BUXTON, P. J., PITKIN, A., & JARVIS, L. J. (2000). The magnetic resonance imaging appearances of the brain in acute carbon monoxide poisoning. *Clinical Radiology*, **55**, 273-280.
- PEARSON, J., BADEN, M. B., & RICHTER, R. W. (1976). Neuronal depletion in the globus pallidus of heroin addicts. *Drug & Alcohol Dependence*, **1**, 349-356.
- QUAMME, J. R., YONELINAS, A. P., WIDAMAN, K. F., KROLL, N. E. A., & SAUVÉ, M. J. (2004). Recall and recognition in mild hypoxia: Using covariance structural modeling to test competing theories of explicit memory. *Neuropsychologia*, **42**, 672-691.
- 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**, 2-13.
- RATCLIFF, R., MCKOON, G., & TINDALL, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 763-785.
- REMPEL-CLOWER, N. L., ZOLA, S. M., SQUIRE, L. R., & AMARAL D. G. (1996). Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. *Journal of Neuroscience*, **16**, 5233-5255.
- ROEDIGER, H. L., III (1990). Implicit memory: Retention without remembering. *American Psychologist*, **45**, 1043-1056.
- RUGG, M. D., & YONELINAS, A. P. (2003). Human recognition memory: A cognitive neuroscience perspective. *Trends in Cognitive Sciences*, **7**, 313-319.
- SMITH, M. L., AUER, R. N., & SIESJO, B. K. (1984). The density and distribution of ischemic brain injury in the rat following 2-10 min of forebrain ischemia. *Acta Neuropathologica*, **64**, 319-332.
- SQUIRE, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 203-231). Cambridge, MA: MIT Press.
- SQUIRE, L. R., & ZOLA, S. M. (1998). Episodic memory, semantic memory, and amnesia. *Hippocampus*, **8**, 205-211.
- TULVING, E. (1985). Memory and consciousness. *Canadian Psychology*, **26**, 1-12.
- VARGHA-KHADEM, F., GADIAN, D. G., WATKINS, K. E., CONNELLY, A., VAN PAESSCHEN, W., & MISHKIN, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, **277**, 376-380.
- WHEELER, M. A., & STUSS, D. T. (2003). Remembering and knowing in patients with frontal lobe injuries. *Cortex*, **39**, 827-846.
- WIXTED, J. T., & SQUIRE, L. R. (2004). Recall and recognition are equally impaired in patients with selective hippocampal damage. *Cognitive, Affective, & Behavioral Neuroscience*, **4**, 58-66.
- YEE, T., GRONNER, A., & KNIGHT, R. T. (1994). CT findings of hypoxic basal ganglia damage. *Southern Medical Journal*, **87**, 624-626.
- YONELINAS, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1341-1354.
- YONELINAS, A. P. (1999). Recognition memory ROCs and the dual-process signal detection model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 514-521.
- YONELINAS, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London: Series B*, **356**, 1363-1374.
- 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., & JACOBY, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory & Language*, **34**, 622-643.
- YONELINAS, A. P., KROLL, N. E. A., 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**, 1236-1241.
- ZOPPELT, D., KOCH, B., SCHWARZ, M., & DAUM, I. (2003). Involvement of the mediodorsal thalamic nucleus in mediating recollection and familiarity. *Neuropsychologia*, **41**, 1160-1170.

APPENDIX

A set of simulations was conducted to determine the likelihood that ceiling effects in recognition scores would produce two memory factors when, in fact, the data were originally generated from a single memory factor. First, it was necessary to derive reasonable parameter estimates of the best single-factor solution on which to base our simulations. A single-factor model was fit to data from the 56 hypoxic patients, where recall and recognition performance was based on a single memory factor (i.e., declarative/recollective memory). We then used estimates of factor loadings (i.e., explained variance) and unique variances (i.e., unexplained variance) to derive population distributions of recall and recognition predicted by the model, and we randomly sampled individual recall and recognition scores for each of the 56 hypoxics in the sample. Ceiling effects were imposed on the distributions of model-generated recognition scores and refit to one- and two-factor models with structural equation modeling to see whether the true factor structure (i.e., one or two factors underlying recognition) would be recovered. We performed the simulation analysis 30 times, each time with a sample size of 56, to get a reasonable estimate of the likelihood that a second factor (i.e., familiarity) would be artifactually produced by ceiling effects in recognition.

For simulating the single-factor model, coma duration and age of each hypoxic case were used to simulate a score on the declarative memory factor for that case. The equation for calculating the memory score was

$$\text{Declarative memory} = b_1 * \text{coma} + b_2 * \text{age} + e_1 * \text{res}_1,$$

where b_1 and b_2 are regression coefficients from the best-fitting single process model predicting the memory factor from age and coma. The individual scores on the coma and age variables were taken from the 56 hypoxic patients, so the means and SDs for these variables were identical to those from the hypoxic sample. The term $e_1 * \text{res}_1$ represents unexplained variance of the memory factor. This term consists of a residual score, res_1 , randomly generated from a normal distribution with a mean of zero and an SD of 1.0, and a weight, e_1 , equal to the square root of the unexplained raw variance of the declarative memory factor. Once a memory score was simulated for each hypoxic, the memory score was used to generate scores on recall and recognition tests from Time 1 and Time 2 (i.e., each subject was tested twice). This involved randomly generating five more scores from normal distributions with means of 0 and SDs of 1.0. Four of these were residuals, one for each test at each time, and one served as a factor score for the recovery factor. The resulting equations for generating recall and recognition scores for each hypoxic are shown below.

$$\begin{aligned} \text{Recall 1} &= b_3 * \text{memory} + e_2 * \text{res}_2 + b_4 * \text{recovery} \\ \text{Recall 2} &= b_3 * \text{memory} + e_2 * \text{res}_3 \\ \text{Recognition 1} &= b_5 * \text{memory} + e_3 * \text{res}_4 + b_4 * \text{recovery} \\ \text{Recognition 2} &= b_5 * \text{memory} + e_3 * \text{res}_5 \end{aligned}$$

where b_3 and b_5 represent the loadings of recall and recognition tests, respectively, on the declarative memory factor, b_4 is the loading of Time 1 tests on the recovery factor, and e_2 and e_3 are the standardized residual SDs of recall and recognition. By using the path coefficients from the best-fitting structural equation model and randomly sampling residuals and factor scores, we were able to preserve the correlational structure of the data expected by the single-factor model and to generate 30 samples of data from this population.

The equations above resulted in mean-deviation scores for recall and recognition in raw score units when age and coma mean-deviation scores are used as input. Then, by adding the mean to each recall and recognition score, the scores were transformed into the metric of the raw observed data. A constant .30 was added to the recognition means to increase the number of simulated recognition scores that fell at or above 15. Ceiling effects were imposed by rescaling all recall and recognition scores falling above 15 and below 0 to 15.0 and 0.0, respectively. The resulting recognition means were highly similar to those based on the real data (means of 11.75 and 12.41 for Times 1 and 2, respectively; 11.59 and 12.45 in the real data). When rounded to the nearest whole number, the average number of subjects at ceiling across 30 simulations was 12.6 participants for recognition at Time 1 (13 for the real data) and 16.5 for recognition at Time 2 (17 for the real data). The number of participants performing at ceiling in the 30 simulations ranged from 8 to 18 for Time 1 and 11 to 23 for Time 2. Thus, we were confident that this procedure effectively reproduced the severity of the recognition ceiling effects seen in the real data.

The one- and two-factor models were fit to the covariance matrices of the simulated data. In the models reported by Yonelinas et al. (2002), we included the constraint that factor loadings for the recall tests be equal across Times 1 and 2, factor loadings for recognition tests be equal across Times 1 and 2, and the recovery loadings be equal for Recall 1 and Recognition 1. When the one- and two-factor models were fit to the 30 simulated samples, the one-factor models converged all 30 times, indicating that the one-factor model was an appropriate model for each of the 30 samples. In contrast, the two-factor model converged only 20 times, suggesting that the two-factor model could not be estimated in one third of the samples. Moreover, in 9 of the 20 samples with converged solutions, we attained convergence only after allowing the familiarity factor to have unequal loadings for the two recognition measures. The ceiling effects of recognition distributions in these 20 samples with converged solutions were as severe as in the larger group of 30 samples (an average of 13.0 ceiling subjects at Time 1 and 16.2 at Time 2). Because the chi-square statistic is additive across independent sam-

APPENDIX (Continued)

ples, we summed the chi-square values for the 20 simulations with converged solutions and tested the significance of this value using the sum of the degrees of freedom across simulated samples. With the simulated data, the overall chi-square value was significant for the 1-factor model [$\chi^2(220, N = 1,120) = 282, p = .003$], indicating that the one-factor model was rejectable statistically. However, the one-factor model was statistically rejectable in only 4 of the 20 samples, indicating that the significant overall chi-square sum likely reflects accumulation of small amounts of misfit present in each sample. Similarly, the overall chi-square was also rejectable for the two-factor model [$\chi^2(191, N = 1,120) = 255, p = .001$], and individual chi squares were significant in 4 of the 20 (all of which also had a significant chi square for the one-factor model), again indicating that the significant overall chi square is likely an effect of adding together small amounts of misfit present in each sample. Although both one- and two-factor models were rejectable using the overall chi-square statistic, this is due to the large overall sample size ($N = 1,120$) across the 20 samples. The chi-square statistic is a function of sample size, so that as sample size increases, the power of the chi-square statistic to reject models is very high even when the models provide close fit to the data.

Importantly, the one- and two-factor models are nested hierarchically such that the improvement of two factors over one factor can be tested statistically by difference in chi square. The resulting sum of differences in chi-square between the models was not significant [$\chi^2(29, N = 1,120) = 27, p = .57$] and is nearly exactly what one would expect by chance alone. Thus, there was no evidence that the second factor added to the model yielded a significant improvement in fit of the model to the data. Furthermore, the two-factor model was a significant improvement over the one-factor model at the .05 level in only 1 of the 20 samples, precisely the rate expected for Type I errors at an alpha level of .05. The simulations support the contention that ceiling effects alone cannot cause false factors to appear in the data any more than would be expected by chance. Furthermore, the structural equation modeling results reported by Yonelinas et al. (2002) and Quamme et al. (2004), which had clear and convincing evidence of a second factor representing familiarity, provide strong evidence for the two-process theory of memory and cannot be attributed to an artifact of ceiling effects in the recognition scores.

(Manuscript received January 29, 2004;
revision accepted for publication July 20, 2004.)