Psychological Review
1993, Vol. 100, No. 1, 3–22

0033-295X/93/$3.00

Novelty Monitoring, Metacognition, and Control in a Composite Holographic Associative Recall Model: Implications for Korsakoff Amnesia

Janet Metcalfe

This article stems from a technical problem in composite-trace distributed models of human memory and particularly in the Composite Holographic Associative Recall Memory (CHARM) model. Briefly, the composite trace—used as a central construct in such models—can become catastrophically out of control. To solve the problem, a prestorage novelty–familiarity *monitor* and a simple *control* procedure need to be implemented. Eight lines of experimental evidence converge on the idea that output from such a novelty–familiarity monitor underlies people's metacognitive judgments of feeling of knowing. Breakdown of the monitoring–control mechanism produces Korsakoff-like symptoms in the model. Impairments in feeling-of-knowing judgments and the failure to release from proactive inhibition, both characteristic of Korsakoff amnesia, are thus attributed to a monitoring–control failure rather than to deficits in the basic memory system.

In his essay "Elbow Room," Dennett (1985) pointed out that the function of monitoring is to permit control. He gave the example of a radio-operated model airplane. To avert a mishap, it is necessary to monitor the plane in its relation to the environment. To make sensible adjustments the operator must know, from moment to moment, the position of the plane, its trajectory and speed, local air currents, gravity, the location of nearby birds and kites, and the positions of trees and must ensure that the model stays within the range of the radio controller. In other systems as well, monitoring is not gratuitous, but rather is an inherent part of a control mechanism necessary for the smooth operation of the system as a whole. The existence of the ability to monitor and predict memory performance (metamemory) is well established. However, its function has received scant attention. There is no clear answer to the question, What is metamemory for? The answer proposed here is that the monitoring system is used to keep the basic memory system under control. It follows that some patterns of memory data may be primarily attributable to the proper functioning of the control system rather than to the basic memory system itself. Hence, certain patterns of results should occur given breakdown of the basic memory system, whereas other patterns are predicted as a function of insult to the control system, and one may delineate the details of this differential breakdown by simulating the model with and without the control system.

I gratefully acknowledge support of this research by National Institute of Mental Health Grant 1 R29 MH48066. I wish to thank Stephan Lewandowsky, Bennett Schwartz, William McGill, Carol Fowler, Thomas Nelson, Laird Cermak, Lynne Reder, Garrison W. Cottrell, Felice Orlich, Noga Gottesfeld, Patricia Churchland, John Welle, and Barbara Dosher for comments and suggestions.

Correspondence concerning this article should be addressed to Janet Metcalfe, Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755.

Results from a large number of studies (e.g., Blake, 1973; Freedman & Landauer, 1966; Gruneberg & Monks, 1974; Hart, 1965; Metcalfe, 1986; Metcalfe & Weibe, 1987; Nelson, Gerler, & Narens, 1984; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Schacter, 1983; Yaniv & Meyer, 1987) have indicated that people have good metamemory. Even though they are unable to recall a particular fact or event, they can assess quite accurately how likely it is that they will be able to recall or recognize it at some later time. This ability to make predictions about future memory holds (though with some variations) across a wide range of tasks, subject populations (A. L. Brown & Lawton, 1977; Cultice, Somerville, & Wellman, 1983; Lachman, Lachman, & Thronesbury, 1979; Wellman, 1977), and even drug states (Darley, Tinklenberg, Roth, Vernon, & Kopelt, 1977; Nelson, McSpadden, Fromme, & Marlatt, 1986). Despite intensive investigation of metamemory capabilities, there has been little mention of the role or function of metacognitions (with the exceptions of Johnson, 1988, 1991; Johnson & Raye, 1981; Nelson & Narens, 1990; Reder, 1987, 1988). In the same spirit as these theorists, I propose that memory metacognitions play a vital role in the smooth and efficient operation of the basic memory system: They are a manifestation of a monitoring and control system that overlooks the basic memory system and is necessary to keep that system from getting out of control.

In the course of so doing, the monitoring–control system functions as an adaptive filter that first assesses the novelty of incoming information and adjusts the sensitivity of the basic memory storage system according to this monitored novelty or familiarity. As I show in this article, some such novelty filtering is a necessary spin-off of a device that keeps the variance of the composite memory trace (a construct proposed in a number of memory models) under control. Thus, the monitoring device is central in allowing the system to be responsive to changing input and to operate dynamically. Schank (1982) and others (e.g., Heath & Fulham, 1988; Levine & Prueitt, 1989) have argued that such adaptive memory storage—dynamically sensitive to the novelty of the to-be-stored events—is necessary for intelligence.

The ideas that (a) a monitoring device is needed to keep the basic memory system from going out of control and (b) such monitoring and control is equivalent to novelty monitoring and filtering stem from studies with a particular composite-trace model of human memory (Metcalfe, 1990, 1991a, 1991b; Metcalfe & Bjork, 1991; Metcalfe & Murdock, 1981; Metcalfe Eich, 1982, 1985). As this article shows, the trace in this model, if left unmonitored and unadjusted, would soon explode out of control. Insofar as the model shares a number of characteristics with human rememberers, it seems plausible to speculate that people have and use a similar monitoring device to circumvent a similar problem. The construct that gives rise to this problem is a central one in this and other distributed models: Memory events are stored in a single composite memory trace.

A number of important psychological implications stem from the notion of composite memory storage. Indeed, this construct is one of the major theoretical and conceptual contributions to the understanding of human memory made by proponents of distributed models (e.g., Anderson, 1977; Anderson & Hinton, 1981; Cavanagh, 1976; Hinton & Anderson, 1981; McClelland & Rumelhart, 1986; Metcalfe, 1990; Metcalfe & Murdock, 1981; Murdock, 1982, 1985; Pike, 1984; Rumelhart & McClelland, 1986). Researchers have argued in favor of this idea on the basis of its neurological plausibility (e.g., Anderson, Silverstein, Ritz, & Jones, 1977; Hebb, 1949; Kohonen, Oja, & Lehtio, 1981), as well as its psychological implications. The most interesting of these include predictions about interference, errors, transformations, prototyping, and distortions that occur in human memory. Despite these attractive properties and predictions, if one assumes that composite-trace models depict real physical systems (see, e.g., Kohonen, 1977), then one must face up to a potentially disastrous consequence of this form of memory storage—under quite ordinary conditions the activation values on the individual elements can increase in magnitude without bounds and out of control. I investigate some implications of implementing a monitoring and control device to stabilize the trace in this article.

I first describe the Composite Holographic Associative Recall Memory (CHARM) model and then show how it suffers from the *exploding variability* problem. Although I use a particular model to illustrate this problem, I believe that the need for and the implications of the monitoring and control device are not specific to this particular memory model but rather apply to a whole class of models. Thus, other models may implicitly make some predictions similar to those of CHARM. I then present simulations to illustrate the exploding variability problem. A solution is outlined that entails adding a simple monitoring and control mechanism to the model. This monitor operating at the time of storage (i.e., preretrieval) amounts to a novelty–familiarity detection mechanism but does not involve retrieval of representational information. As such, it can be used for making metacognitive judgments, particularly those that involve feeling-of-knowing judgments. Eight lines of experimental evidence are surveyed that point to such a preretrieval locus for these judgments.[1]

Finally, if there exists such an adjunct monitoring and control mechanism, then one might expect to find cases in which it is disrupted. Korsakoff patients appear to suffer from such a breakdown. As is discussed, Korsakoff patients show selective impairment in their judgments of feeling of knowing (Janowsky, Shimamura, & Squire, 1989; Shimamura & Squire, 1986). In addition, a hallmark of Korsakoff amnesics is that (unlike other kinds of amnesics) they fail to release from proactive inhibition (Moscovitch, 1982; Squire, 1982). Computer simulations are presented to illustrate the behavior of the model—with and without the control system—in a release from proactive inhibition paradigm. The results of these simulations are remarkable: With the monitoring and control system in place, the model releases from proactive inhibition given category shifts, much in the manner of "normal" subjects. Without the monitoring and control device, however, the model produces results typical of Korsakoff patients, specifically, a failure to release from proactive inhibition.

Description of the Model

The CHARM model incorporates the idea that the results of many associations or events are stored by being superimposed in a composite memory trace. Because of this superposition the elements necessarily interact (see Metcalfe, 1990; Metcalfe Eich, 1982, 1985; Murdock, 1982). It is the nature of this composite trace that generates the problem that is of focal interest here (and provides the opportunity for the solution that is explored).

Representation

Items in the model are represented as vectors with values randomly distributed around zero. They vary in their similarity to one another, as prescribed by the experimental situation. The model allows for more specific delimitation of the exact makeup of particular items if the experimental situation or the nature of the items themselves warrant it. As is shown, different kinds of items, particularly items that bear different kinds of relations to the information already stored in the composite memory trace, have differential impacts on the variance of the trace. Unrelated events increase the variance to a minimal extent, whereas highly related items have a greater effect.

Association Formation

Two items, $\mathbf{A} = (a_{-(n-1)/2}, \ldots, a_{-1}, a_0, a_1, \ldots, a_{(n-1)/2})$, and $\mathbf{B} = (b_{-(n-1)/2}, \ldots, b_{-1}, b_0, b_1, \ldots, b_{(n-1)/2})$, are associated in the CHARM model by the operation of convolution, denoted $*$, and defined as:

[1] Many other models use feedback to adjust the trace as a function of what is retrieved. Those models do not make the same predictions as CHARM, in which the control mechanism is not related to the retrieval mechanism. Thus, although CHARM makes the prediction that feeling-of-knowing judgments and other monitoring–control-related judgments and patterns of data may be dissociable from memory retrieval, the retrieval-based feedback models do not allow such dissociations.

$$(\mathbf{A} * \mathbf{B})_m = \mathbf{T}_m = \sum_{(i,j) \in S(m)} a_i b_j,$$

where $S(m) = \{(i, j) \mid -(n-1)/2 \leq i, j \leq (n-1)/2, \text{and } i + j = m\}$. The subscript m denotes the mth element in the vector formed by convolution. Interitem convolution primarily underlies recall, whereas autoconvolution primarily underlies recognition.

Storage

The results of successive convolutions (whether they are autoassociations or interitem associations) are added into a single composite memory trace. The trace, **T**, consists of associations as follows:

$$\mathbf{T} = \mathring{a}\mathbf{A} * \mathbf{A} + \beta\mathbf{A} * \mathbf{B} + \mathring{a}\mathbf{B} * \mathbf{B} + \mathring{a}\mathbf{C} * \mathbf{C} + \beta\mathbf{C} * \mathbf{D} + \mathring{a}\mathbf{D} * \mathbf{D} + \ldots + \text{preexisting noise}.$$

The weightings for the autoassociations (å) and for the interitem associations (β) may be varied according to the experimental situation. The trace is also assumed to start out with some "noise" from previous memories.

Retrieval

The retrieval operation is correlation (see Uttal, 1975, 1988, for some other examples of this operation in human cognition–perception). Retrieval generates a new vector $\mathbf{R}_m$ from the elements of the cue and trace vectors by cross-correlating them. Accordingly,

$$\mathbf{R}_m = \sum_{(i,j) \in S(m)} q_i t_j,$$

where **Q** is the cue vector with elements q_i, **T** is the trace with elements t_i, and $S(m)$ is the domain of paired elements over which the correlation is attempted. That is, $S(m) = \{(i, j) \mid -(n-1)/2 \leq i, j < (n-1)/2, \text{and } i - j = m\}$. The result of retrieval is a new vector reflecting what the subject generates from episodic memory. Unlike the case in many network models, the CHARM model is not taught what to retrieve; it simply produces, as the retrieved item, whatever results from the correlation process. This item may be noisy or systematically distorted. The task is then to find whether the distortions, errors, and correlations produced by the model match those produced by human subjects.

Recall

Recall is based on retrieval, that is, on the vector resulting from correlation of the cue with the trace. However, because the output vector is typically noisy or distorted and because in simulations of the model it is necessary to say what is recalled and how frequently, a decision process is also necessary. The decision process is formulated as follows: The retrieved vector is matched to every item in a lexicon of possible outcomes, and, in the simplest case, the item yielding the highest dot product will be the item recalled. However, this dot product must exceed a lower threshold. If the retrieved signal is too noisy to be interpretable, recall will not occur. This threshold on recall controls the intrusion rate.

There is no a priori reason that the output from the retrieval process could not be fed directly into a motor program instead of a lexicon if the output required were a motor response rather than a word. For example, if the input to the episodic system had been vectorized pictorial images that could be reconverted by way of some nonlexical system into their pictorial form, then the output from the retrieval process in the holographic model could also be fed directly into this converter and displayed as pictures. My colleagues and I (e.g., Cottrell & Metcalfe, 1991) are working on such a system, using a back-propagation image-compression network as the preprocessor. When the materials are verbal, I shall use a localist lexicon for the decision process. Although there may be a discrete localist representation for words and perhaps for other mental objects,[2] commitment to a lexical representation for every kind of input to memory is neither necessary for the CHARM model nor necessarily implied by its use with this one form of input.

Recognition

Recognition is also based on retrieval. In this case, however, the result is matched only against the probe itself. If the probe was autoassociated and entered into the composite trace, then this would show a positive dot product (or resonance value) with the probe. If it was not encoded (and was unrelated to everything that was encoded), then the match would have an expected value of zero. A *yes* decision would be given if this match was to exceed a particular criterion.[3]

---

[2] Early distributed models held to the view, based on the equipotentiality studies of Lashley (1950), that the human cognitive-perceptual system must be distributed throughout. By this view, a localist lexicon seems contraindicated. However, recent neuropsychological evidence (see, e.g., Farah, 1990; Shallice, 1988) has pointed toward distinct systems or modules, some of which are distributed (like the association and composite trace in CHARM) and some of which are symbolic and localist (like the lexicon in CHARM). The contrast and interplay of the distinctive characteristics of these two kinds of systems is a most exciting research area.

[3] The issue of whether models such as CHARM can learn and whether they do or do not suffer from catastrophic interference has some currency. In a previous study (Metcalfe Eich, 1982), I showed by computer simulation and discussed how the model learns in the classic A–B A–C paradigm on which back-propagation models fail to show humanlike patterns. CHARM automatically generated the pattern produced by subjects (a nice trade-off between B and C responses and no catastrophic interference), and it also exhibited independence between the B and C responses, as do people. Another set of simulations modeled the Osgood-Martin surface, in which the similarities of the cues and targets were systematically varied. The model again produced the same pattern of results as people. There was one snag (as was discussed in that article): If the memory trace in the model started as a zero vector—an implausible assumption, but one sometimes made for mathematical convenience—and the exact pair was repeated and not contrasted to anything else, both the signal and the noise increased at the same rate, and no ostensible learning took place. A more plausible assumption is that the trace starts out as noise rather than as a pristine blank slate. When the trace started out as random noise—which could be due either to noise in the nervous system or to preexperimental events not under focal consideration in the experiment—the model produced the classic human learning results without further ado.

The Problem of Exploding Variability

As noted earlier, to construct the composite memory trace, successive random (or sometimes nonrandom) vectors are added. With an increasing number of additions of vectors into the composite trace, the variance computed over the elements of that trace increases. The situation is described by Feller (1957): "If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are random variables with finite variances of $\sigma_1^2, \ldots, \sigma_n^2$, and $\mathbf{S}_n = \mathbf{X}_1 + \ldots \mathbf{X}_n$, then

$$\mathrm{Var}(\mathbf{S}_n) = \sum_{k=1}^{n} \sigma_k^2 + 2 \sum_{j,k} \mathrm{Cov}\,(\mathbf{X}_j, \mathbf{X}_k),$$

the last sum extending over each of the $\binom{n}{2}$ pairs $(\mathbf{X}_j, \mathbf{X}_k)$ with $j < k$" (p. 216). S_n in Feller's equation is the trace in the model, and the **X**s are the association vectors being added into that trace. If the $\mathbf{X}_j$s (or the incoming associations) are independent of one another, then the covariance terms drop out and the addition rule applies, resulting in a linear increase in variance as a function of the number of associations being added into the trace: Var $(S_n) = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$. Doubling the number of random vectors doubles the variance of the trace. When the vectors being added to one another are similar, the problem is much worse because of the covariance terms. These results are illustrated with simulations.

The problem of increasing trace variability was first pointed out in an early version of a convolution–correlation model (Metcalfe & Murdock, 1981) that stood halfway between the seminal archetype model of content addressable distributed associative memory (CADAM; Liepa, 1977) and the more detailed, refined, and psychologically plausible descendents that are being explored today—Theory of Distributed Associative Memory (TODAM; Murdock, 1982, 1985), Lewandowsky and Murdock's closed-loop and open-loop models (1989a and 1989b, respectively), Weber's (1988) redintegrative model, Hockley and Murdock's (1988) decision model, and the CHARM model. To my knowledge, despite advances in other domains, none of the aforementioned models have solved the problem.

Simulations

To illustrate the problem in a more concrete form, I ran several simulations of the CHARM model and printed the memory vectors as bar graphs (such that each element of the vector corresponded to a bar). The height of each bar gives the activation of that element. Values can be either positive or negative, and one might wish to think of these values as being either excitation or inhibition (or potentiation or inhibition) of firing on neurons, instead of, or in addition to, more psychological meanings that one can ascribe to the patterns. In the first case, the items in the lexicon were constructed to be all unrelated to one another; that is, each was a random vector with respect to all of the other vectors, and each was normalized so that the lengths were one. The items consisted of 31 features, and the composite memory traces illustrated in Figure 1 were truncated to the central 31 features. The top panel shows the trace with just one pair, the middle panel with three unrelated pairs, and the bottom panel with five unrelated pairs. Although the values on the trace remained scattered around zero, the variability increased.

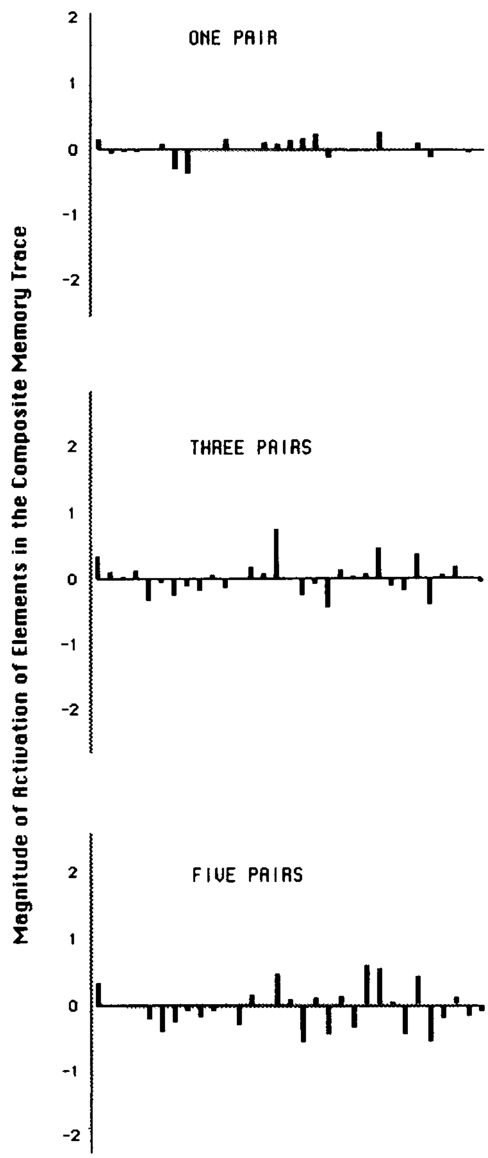

Figure 1. Composite memory trace with increasing number of unrelated pairs.

Of course, it is unknown what the scale of the memory storage system might be or what its range of sensitivity is. Nevertheless, assuming that this trend continues, if this is a real physical system, at some point the trace will be pushed beyond its capabilities. In particular, if the elements of the model correspond to neurons, the potentially boundless increase in the activation on the elements is unacceptable. I later assume that the system is kept under control on a strict item-by-item basis. The default that is taken, then, is that the variance of the trace is fixed rather than potentially increasing.[4] Although some control is

[4] The idea that the variance is fixed is a working first approximation and is probably too simple. Given that the variance of the trace reflects the capacity of the episodic memory system at time t, then it may be

necessary, the particular kind of control and strictness of control was chosen for its simplicity, and other possibilities deserve investigation as well.[5]

Figure 2 illustrates the trace when the items were related to one another. To make the items related, the lexicon was first set up to consist of unrelated items. Then an item was arbitrarily chosen to be a category prototype. Category exemplars were constructed by randomly selecting a subset of features on each exemplar and replacing the random values on those features with the values that had been assigned to the category prototype. The tightness of the category was manipulated by varying how many such features were selected: A large proportion produces a tight or highly structured category; a small proportion produces a loose category structure. As the three panels of Figure 2 show, the trace variance increases more with the similar materials than it did with the unrelated materials.

Figure 3 takes the extension of similarity to an extreme. Identical pairs were convolved and added into the trace. It is easy to understand why the trace expands in this maximum similarity case: The trace with five associations is simply the trace with one association multiplied by five.

To summarize the results shown in Figures 1 to 3, Figure 4 gives the average square roots of the sums of squares that were computed for the traces shown in Figures 1 to 3 plus the intervening traces that were simulated but not shown in the preceding figures. Two hundred replications of each of the above simulations (and simulations in which the number of pairs entered into the traces were two and four) were conducted, and the average measures of the variability of the traces were computed over the 200 runs. The relation between these variabilities and the similarities of the items in the traces is clear from Figure 4. The increase in variability is directly related to the similarity between the items entering into the trace and the trace.

Simulations With Category Groupings

I conducted a number of simulations of a situation much like that given in Figure 2 except that multiple categories were constructed. The trace consisted successively of up to 5 pairs all from the same category. Then, for the next 5 pairs, the category was switched to one that was unrelated to the first. I extended the simulation up to 25 such pairs, computing the sum of squares of the trace as each new pair was entered in. The results presented in Figure 5 are averaged over 200 independent runs of the simulation. As Figure 5 illustrates, the variability increased precipitously within category, but then less steeply when a shift was made. As more and more similar pairs were entered into the composite trace, the slope of the variability accelerated. Overall, a scalloped pattern is produced from category shifting. This relates to data, in the release from proactive inhibition paradigm, that are of importance in distinguishing Korsakoff amnesia from other kinds of amnesias.

Solution to the Problem: Monitoring and Control

The results of the first simulation with unrelated items seemed simple enough. The increase in variability is linear. If that were all that was going on (and one could depend on the entries to the memory trace to be unrelated to one another and to everything else in the trace as well as to be occurring at a constant rate and with the same weighting or attentional importance), one could solve the problem with just the right decay parameter. Such a fixed-parameter solution does not require a monitoring device. However, such a solution will not work.

Monitoring is needed because the nature of the to-be-stored event and its relation to the trace may vary. What new to-be-remembered events might occur in the stream of one's experiences are not known, and one certainly does not know that such events will all be unrelated to everything else that has come before. Insofar as these variations *matter*, as was shown in the foregoing simulations, it must be determined how any particular entry will affect the variance of the trace. The effect depends on the nature of the event itself. Are the two items of the pair unrelated or similar to one another, and, if the latter, how similar? The trace variance also depends on the relation of the to-be-stored event to the trace. Is the event unrelated to anything else that is stored in the trace, or is it correlated with the trace? If the latter, how correlated?

It seems unlikely—indeed, impossible—that the human memory system would solve this problem analytically. However, computationally, only one step is needed for monitoring and another step for control. The variance that would result were the association entered directly into the trace needs to be assessed. This is done by the monitor. In the simulations that

desirable to allow that attentional and arousal factors alter this capacity on a moment-to-moment basis. It may tend to be a bit higher under conditions of high arousal and lower than normal under conditions of stupor. Study of the implications of this construct could be most instructive, but I have delayed their exploration within the context of CHARM until the basic implications of the construct of control have been mapped out.

[5] A particular (simple) method of renormalization was adopted here, for controlling the variance. Other methods exist. For example, the trace could decay and the new incoming event alone could be weighted inversely according to its novelty. Alternatively, the values on the elements of the trace could be truncated at some threshold. Truncation produces an information-loss problem and eventually allows no further information storage, thus some form of renormalization (rather than simple truncation) seems most reasonable for the first approximation to what is needed in the episodic memory system, which must remain plastic. Differences between the types of renormalization are unimportant for the major points addressed in this article. Both methods give rise to a monotonic difference in the monitored value of the increase in variance attributable to the incoming event as a function of event-trace congruence (though the numerical values for the two schemes are different). Both occur prior to retrieval and hence could underlie feeling-of-knowing judgments. Both would give rise to the release from proactive inhibition phenomenon. They are testably different insofar as the simple renormalization procedure makes the prediction that highly familiar incoming events decrease the memorability of everything that came before. This prediction is not made by the more complex mechanism that specifies trace decay plus selective novelty filtering only on the incoming event. Further experimentation is needed to test these predictions.

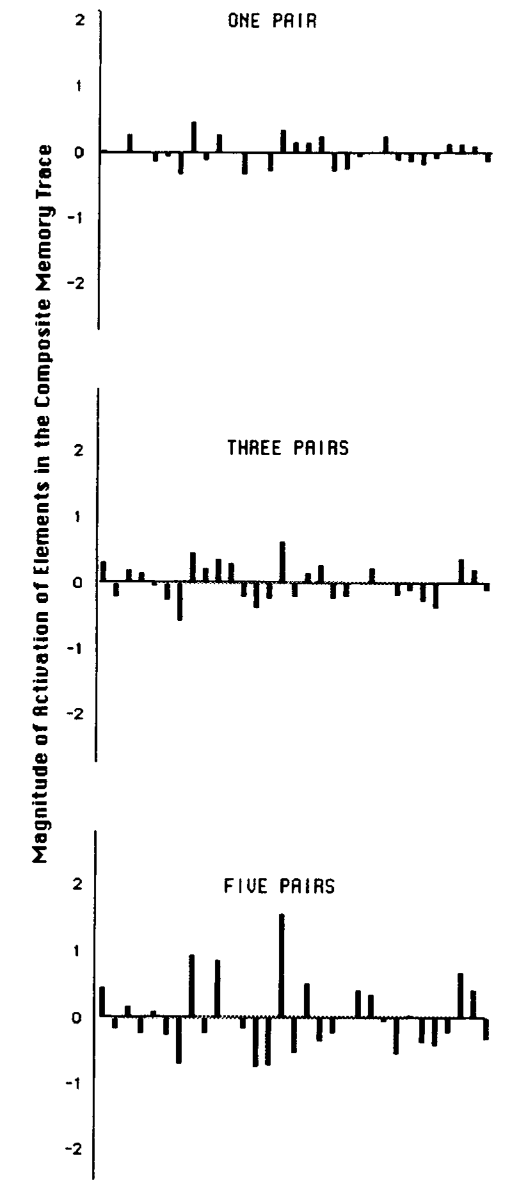

Figure 2. Composite memory trace with increasing number of similar pairs.

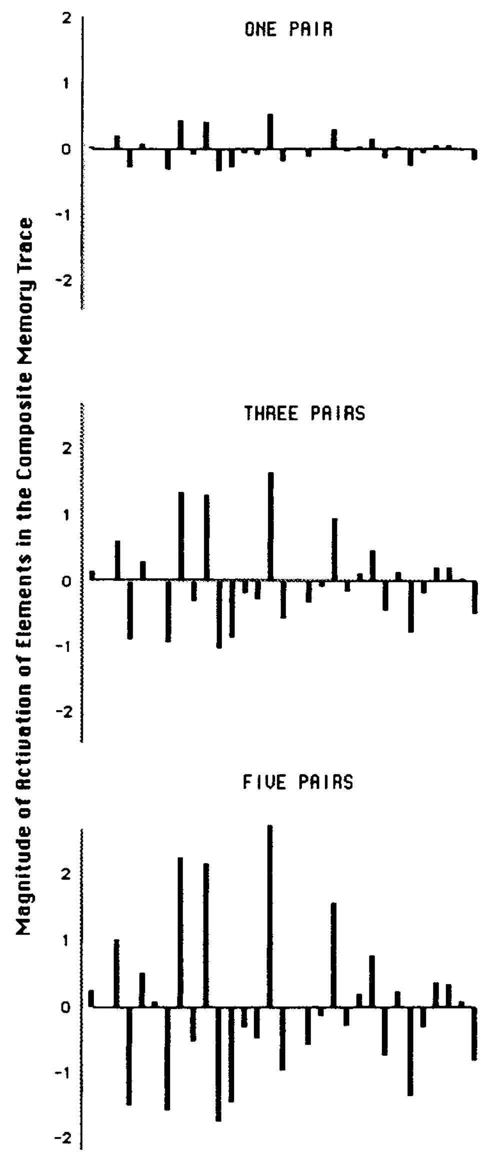

Figure 3. Composite memory trace with increasing number of identical pairs.

follow, a familiarity value, the sum of squares, is computed as a measure of this variability (i.e., $F = \Sigma [T_m + (A*\mathbf{B})_m]^2$, where F is the measure of familiarity, **T** is the old trace, and **A*B** is the newly convolved vector being added into the trace). Then the result of this computation is used to weight (control) the combined new trace: It is simply used as the denominator in renor malization, so $T_{new} = 1/F\ (\mathbf{T}_{old} + \mathbf{A*B})$. The new trace is thus weighted such that the old trace variance equals the new trace variance.

If the new event is highly similar to what is in the trace, the monitored variance will be large and the denominator will be large, which means that the weighting on the new entry will be small. If the new entry is unrelated to the old trace, the new variance will be relatively small. This means that the denominator for renormalization will be small and the weighting on the newly entered event will be large.

Novelty-Familiarity Monitoring and Filtering

These two simple processes are necessary to keep the variance of the trace from exploding. The variability that the monitor computes (which is a scalar) is actually a measure of novelty or familiarity. A judgment based on this value could be made quickly because neither specific retrieval of an item (a vector) nor the decision processes attendant on retrieval need occur. Rather, such a judgment would only involve the first step in entering a new event into memory—a normal part of what must be done in any case. This value does not contain any specific information such as characteristics of the stored items, how-

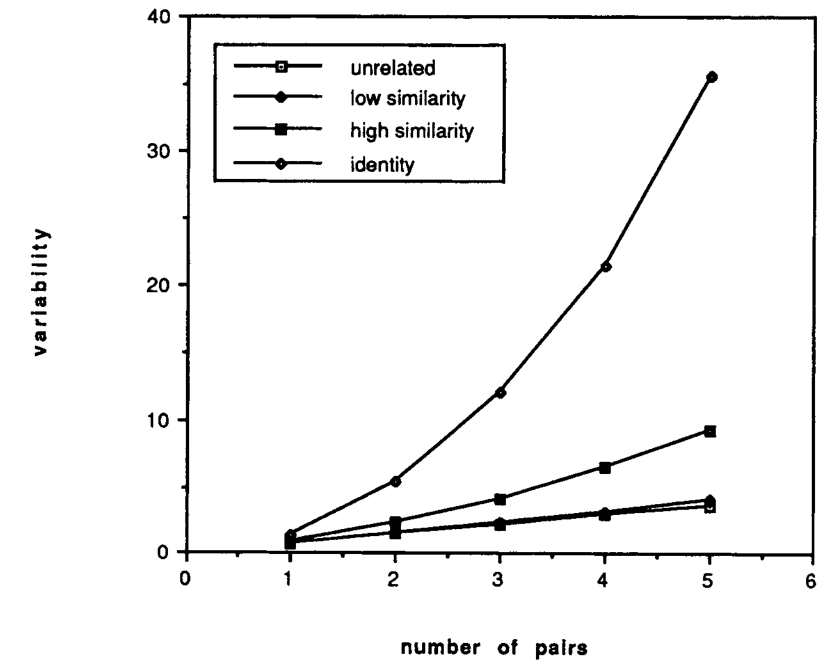

Figure 4. Variability of traces as a function of the number of pairs and the similarity of the pairs.

ever.[6] If the value returned by this computation is large, it means that the event is highly familiar. If the value the monitor returned is very low, the event is novel.

Adjusting the trace as a function of this monitored value could be called familiarity–novelty filtering. If the event-trace combination returns a large variance, it is assigned a low weighting during the renormalization process to keep the variance stable. One could call this weighting allocation of *cognitive energy* (see Kahneman, 1973, for a discussion of related ideas in effort-based attentional research). It follows that old and familiar events are given little cognitive energy. Habituation results. If the event is novel, it can be given a high weighting. Novel events, then, get extra cognitive energy (i.e., they are arousing; the person is alerted to novelty; or novel events receive more attention).

Trace Simulations With Monitoring and Control

The results shown in Figure 6 are from a sequence of simulations in which this renormalization process took place with every new entry into the memory trace. The list structure was the same as in Figure 5. Several categories were constructed that were unrelated to one another but in which the within-category exemplars were highly similar to one another and to a category prototype. First, five pairs from one category were added to the composite trace. Then there was a shift to a new category for the next five pairs. In the present simulations the trace was renormalized each time a new association was entered into the trace. First, the sum of squares was computed with each new entry as shown in Figure 5. Hence, the figure shows what the monitoring device "sees." Then this value was used to renormalize the trace so that the variance of the trace was squeezed down to 1.0 (not shown in the figure). When the next new input came in, the new variability was computed (shown in Figure 5), and then the trace was again squeezed down to have a variance of 1.0.

Figure 6 illustrates that the first pair of a category increased the variance the least. The next pair within the category increased the variance more, even though the entire trace had been squeezed down to a variance of 1.0 immediately before its entry into the trace. Because of the contribution of the first pair to the trace, the second pair was more correlated to the trace when it was presented than was the first pair when it was presented. Each successive entry produced a more and more variable trace. This trend continued until the category shifted at the fifth pair. Because the first pair in the new category was unrelated to everything in the trace, it contributed little to the variance. The weightings assigned at each step are inversely proportional to the variability measures shown in Figure 6.

As mentioned earlier, the monitor that computes the variance ascertains novelty–familiarity and could be used as a

[6] Because the value returned by the monitor is just a scalar and contains no representational information, it could not be used for implicit memory tasks such as fragment completion or anagram solving, which require representational information about specific items. Although some researchers have used the term *familiarity* (possibly with a different referent in mind than the scalar value specified in the model) to explain improvements in these tasks, the scalar familiarity value returned by the monitor does not provide a sufficient condition to account for how people are better able to solve fragments or anagrams given prior exposure. As outlined in the Priming of the Targets section, priming effects in CHARM are attributable to lexical activation.

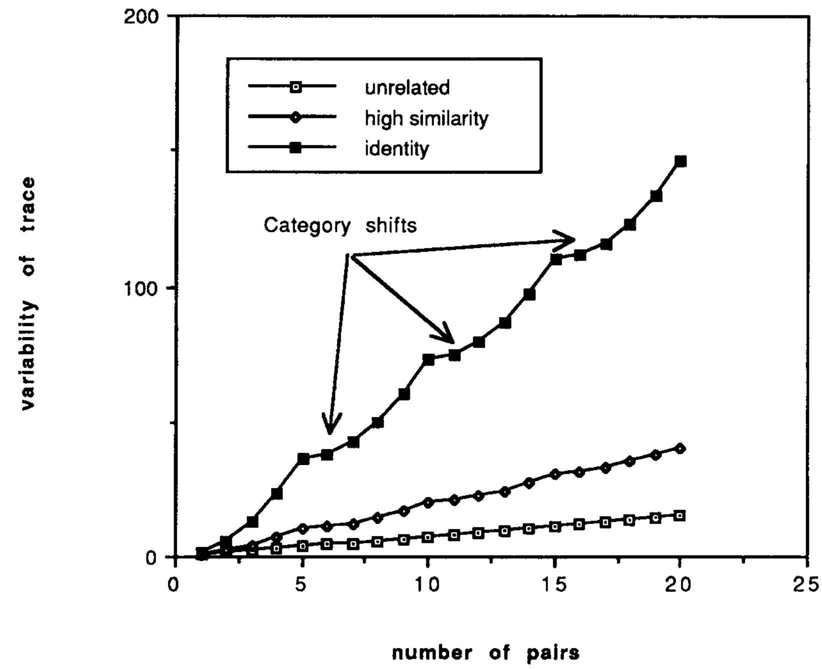

Figure 5. Variability of traces as a function of category similarity. (The category changes at every fifth entry, as shown by the arrows.)

quick indicator of feelings of familiarity. If the computed variance is high, it means that the item about to be entered into the trace is familiar; if it is low, it means that the event is new. Some evidence that people may have access to such a fast familiarity measure (which is not based on the explicit retrieval of representational information) and that they can use it for making metacognitive judgments of their feelings of knowing, is reviewed in the next section.

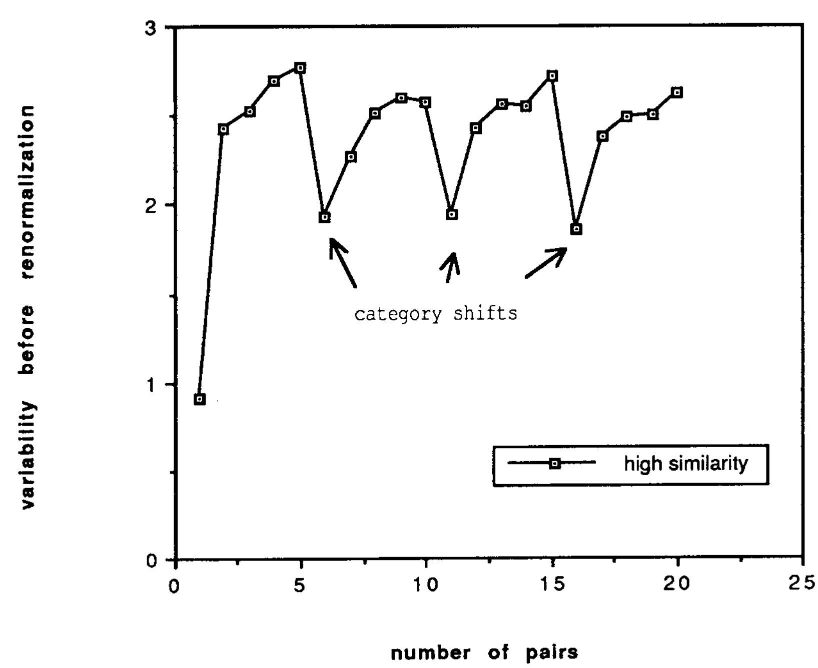

Figure 6. Variability as a function of category similarity when the trace is renormalized after every entry. (This figure shows what the monitor "sees" with each new input. There is a category shift with every fifth entry, as shown by the arrows.)

Feeling-of-Knowing Judgments

When considering the basis for people's feeling-of-knowing judgments, probably the most intuitive conjecture is that these judgments result from retrieval of the to-be-remembered information (see Koriat, 1991, for elaboration of this view). Although I do not wish to deny the possibility that some judgments might indeed be based on retrieval of partial information, there exist considerable data indicating that this cannot be the whole story and may not even be the main story. These data, outlined later, suggest that feeling-of-knowing judgments can be made on the basis of information resembling that gleaned by the novelty–familiarity monitor in the model. To allow judgments made on this basis (in the model), the cue would be associated with itself and treated as if it were to be entered into the trace. Note that in the most current version of the CHARM model (Metcalfe, 1991b), as in earlier versions concerned with recognition (Metcalfe Eich, 1985), the cue is associated with itself, the cue and the target are associated, and the target is associated with itself. The autoassociations are primarily responsible for recognition memory and the interitem associations are mainly responsible for recall. However, because of the composite nature of the trace and the interactive nature of the associations, there is spillover between and among associations. So the assumption here that the cue is autoassociated is only to say that the model processes the information in its usual way.

The novelty monitor then also functions in the usual way—assessing the novelty of the to-be-stored incoming information. If the cue were highly familiar (i.e., if it were related to events already stored in the trace), it would increase the variance by a considerable amount, and the monitor would return a high value. The person could use this high monitored value to make a *high* feeling-of-knowing judgment. If the cue did little to the variance of the trace, then the subject could use this low monitored value to make a *low* feeling-of-knowing judgment. In short, this computation, which needs to be done routinely, could be used directly to make metacognitive judgments. Eight converging lines of evidence supporting this idea are reviewed in the next section.

The Standard Paradigm for Feeling-of-Knowing Judgments

There have been a large number of feeling-of-knowing studies. The task has been studied with different groups of subjects ranging from very young to quite old, with different kinds of materials, with differing amounts of learning, and so on (see Nelson, 1988, for a review). In many of these studies, subjects are initially asked to recall the answer to a question. For example, subjects might be asked "What is the capital of Canada?" or "Who was the first prime minister of Canada?" or "Who is the current prime minister of Canada?" The feeling-of-knowing procedure, whereby subjects either rank order or give estimates of the probability that they will later be able to remember the answers, is typically enacted *on only those items that the subject cannot remember* at the time of initial test. Feeling-of-knowing judgments for this irretrievable information, as compared to later recognition of the information are quite accurate, with correlations usually between .45 and .55. Because, by definition, subjects cannot retrieve the explicit information and yet can make accurate judgments, it is plausible to suppose that the judgments are made on the basis of information other than that which is explicitly retrieved.

Differences in Feeling of Knowing Among Different Error Types

As mentioned earlier, the feeling-of-knowing task focuses on those memory events that subjects cannot initially recall. Such nonrecalled answers can be divided into two classes: (a) errors of omission, where nothing is recalled, and (b) errors of commission, where the wrong answer is given. Presumably, if one were basing one's feeling-of-knowing judgment on what was retrieved but knew that what had just been retrieved was wrong, the feeling-of-knowing rating for that item should plummet. Errors of commission might be expected to produce especially low ratings.

On the other hand, if the judgments were based on the value returned by the novelty–familiarity monitor, predictions about the ranking of commission and omission errors are different. The familiarity measure returned by the cues producing commission errors should be high. These cues were familiar enough to retrieve something from memory. The cues for omission errors were so unfamiliar that they produced no response. Krinsky and Nelson (1985) have analyzed feeling-of-knowing ratings as a function of these two error types. They found that the ratings given to errors of commission were much higher than the ratings given to omission errors, which is consistent with the familiarity–novelty monitoring hypothesis of feeling of knowing.

Link to Study Time

Feeling-of-knowing ratings are related to subsequent self-paced study time (Nelson & Leonesio, 1988). This finding makes sense in terms of the role of the monitoring device in the eventual control of the weighting of various items that are entered into the trace. Study time may be one way in which the appropriate weightings are actualized—events contributing little variance receive a higher weighting (study time) than those contributing a large amount of variance.

To give a slightly more concrete example of how this linkage might work, one might ask how, physically, division (or multiplication) is accomplished in the nervous system. One possibility is that the multiplicand acts as a timing gate that allows the stream of incoming information to have its effects for varying amounts of time. This idea is common in neuroscience. If the number by which the trace is multiplied is small, the information that is being added into the system would have its effect for only a short amount of time; if the number is large, then the effect would be for a longer amount of time. Thus, the link to measured study time is straightforward. There may, of course, be other ways in which the translation between monitored value and time are connected.

Feeling-of-Knowing Latency Compared With Retrieval Latency

If feeling-of-knowing judgments could only be based on explicitly retrieved information (but in the special case where there is insufficient information to allow a discrete response), then one might expect that latencies to make these judgments would be slower than retrieval latencies. In two studies investigating judgment versus retrieval latencies, Reder (1987, 1988) used a technique dubbed the "game show" paradigm. Subjects in one group were asked to say *yes* or *no*, indicating that they knew they could provide the answers to general information questions. In the other group, subjects were to provide the answers. The dependent measure was the time to initiate a response. A second experiment used button presses in both cases, to control for the possibility that verbal differences across conditions might have had some impact on the results. Thus, subjects hit the *yes* button when they had the answer (in the retrieval group) or when they knew they would be able to get the answer (in the feeling-of-knowing group).

In both experiments, the responses were faster in the feeling-of-knowing condition than in the retrieval condition. Reder's (1987, 1988) findings provide support for the preretrieval locus of feeling-of-knowing judgments. She suggested that fast feeling-of-knowing judgments might provide the basis for deciding whether to initiate retrieval (or some other question-answering strategy) at all. In the present context, this finding of a faster latency to make these metacognitive judgments than to begin to initiate a response indicates that the feelings of knowing are not just derivative from the response, but may be due to a separable (early) process.

Priming (or Familiarity) of the Cues

If feeling-of-knowing judgments were based on a novelty monitor that assesses the familiarity of the cue, then manipulations that increase cue familiarity should increase feeling-of-knowing ratings. This result should be obtained regardless of whether the manipulation increased the likelihood of retrieval of the target information. In contrast, if feeling-of-knowing judgments were based on retrieval of partial target information, then manipulations that affect only the cue familiarity but have no effect on recall should not change subjects' feelings of knowing.

Reder (1988) devised a method for altering only the cue familiarity. Before being given general information questions in a feeling-of-knowing task, subjects (in an ostensibly unrelated task) rated a list of words for frequency of occurrence. Embedded in the frequency-rating task were some of the words that occurred later in one third of the feeling-of-knowing questions. For example, the words *golf, par, clown,* and *Howdy Doody* might have been presented on the frequency rating list. In the feeling-of-knowing task cues appeared, such as "What is the golf term for a score of one under par?" or "What was the name of the clown on the Howdy Doody television show?", as well as cues that were not primed. The results showed that priming the cues spuriously increased subjects' feelings of knowing without increasing their ability to answer the questions.

Reder (1988) interpreted her data as indicating that certain kinds of metacognitive judgments, such as feeling-of-knowing judgments, may be made very quickly, and they are not based on explicit retrieval—a conclusion with which I concur. These judgments, she argued, can then be used by subjects as the basis for deciding what kind of memory or problem-solving strategy to use. She presented a number of other lines of argument to support these ideas, not only for the preretrieval locus of metacognitions but also for their control function. The results of the cue-priming study provide good support for an early, familiarity-based locus for feeling-of-knowing judgments, as opposed to a partial-information retrieval-based locus.

Converging evidence that the familiarity of the cues, rather than target-specific information, underlies feeling-of-knowing judgments comes from a sequence of experiments by Glenberg, Sanocki, Epstein, and Morris (1987). In a typical experiment, subjects were given 15 short informative paragraphs to read, each on different topics. They were then given the titles of the stories and asked to predict (i.e., give confidence ratings about) either their specific recall of aspects of the stories' content or to predict their ability to make appropriate inferences about each story. The inferences did not necessarily tap domain knowledge but rather were specific to the paragraphs read during the experiment.

The basic findings were that subjects' predictive accuracy in these experiments was near zero. Interestingly, though, there was a large positive correlation between the subjects' domain knowledge, or the familiarity of the cues, and the subjects' confidence ratings. Glenberg et al. (1987) interpreted these findings as indicating that judgments were based on domain familiarity of the cue. As in Reder's (1988) experiments, then, these experiments suggest that feeling-of-knowing assessments are based more on assessment of the cue than on attempted partial retrieval of the target. In many cases, of course, familiarity of the cue and retrieval of the target are well correlated. Typically, the correlations between feeling of knowing and later recall performance are high. However, when the correlation between cue familiarity and target retrieval are teased apart, the judgment appears to depend more on the cue familiarity. Target information appears to have relatively little impact on the judgment, consistent with the idea that subjects use the kind of familiarity monitor delineated earlier as the basis for feeling-of-knowing judgments.

Priming of the Targets

As noted in the foregoing section, priming of the cues (Reder, 1988) resulted in spurious increases in the feeling of knowing, without a concomitant increase in recall. The experiments described by Jameson, Narens, Goldfarb, and Nelson (1990) provide the converse result. In their experiments the targets were primed. Although this priming resulted in an increase in recall it produced no discernible effect on feeling-of-knowing judgments.

Jameson et al. (1990) reported two experiments in which the answers to general information questions were presented to subjects at a rate that was at or close to the threshold of consciousness (using a variant of Marcel's 1983 procedure). Subjects were given a series of general information questions in a pretest to determine those unrecallable answers that would thereby qualify for the feeling-of-knowing test. Subjects' individual near-threshold durations were also determined by a pretest. Then, for example, subjects were presented with a question such as "What is the name of the North Star?," which had been a question on a pretest that the subject had been unable to answer. Following cue presentation they were flashed either the answer (Polaris) or a nonsense word that followed the statistics of English words. In both conditions this was followed by a mask, at a timing determined by the pretest to minimize the chance that the subject was consciously aware of the flashed item. Recall and feeling-of-knowing judgments followed the masked presentation of the target or the nonsense word.

There were two basic findings. First, recall was enhanced by the near-subliminal presentation of the target. When the targets had been flashed, recall was 18% higher than when the nonsense words had been shown. Second, the manipulation had no effect on feeling-of-knowing ratings. Indeed, in one experiment the feeling-of-knowing correlation was higher in the nonsense-word condition. In both experiments, the absolute magnitude of the feeling-of-knowing ratings was unchanged by flashing the answers nearly subliminally.

In terms of the model, one would expect the priming manipulation in this experiment to have increased the activation level of the target items in the lexicon. However, given that this information may not have been conscious, and consciousness is a prerequisite for entry into the episodic or explicit (hippocampal) memory system, the flashed targets may not have entered the episodic composite trace at all. If the subliminal information did not enter the composite trace, it could not have an impact on judgments based on the trace alone. Even if the information about the target had entered into the trace, unless it was associated with the cue, it would not necessarily increase the familiarity of the cue to the trace. The lack of effect on feeling-of-knowing judgments is thus explicable. An explanation, however, is needed about why recall was positively affected by the priming manipulation.

To take the most extreme case, assume that the primes only activated items in the lexicon, without ever entering into the composite trace. Such lexical priming may nevertheless influence cued recall because of the importance of the lexical activation values in the decision stage of recall. Suppose that the item retrieved from the composite trace to a cue like "What is the name of the city that houses the U.S. Naval Academy?" is noisy. Under unprimed conditions, it resonates (i.e., has a positive dot product) with a number of lexical entries. Recall that the lexical item having the highest resonance with the retrieved vector is the response that is given as the recalled item, as long as that resonance value exceeds a criterion. The noisy retrieved item might thus evoke the wrong response or, if none of the resonances are great enough to exceed criterion, no response. But suppose, now, that the correct lexical item with which the retrieved item would resonate to some extent in any case (but perhaps not enough) has been strengthened or primed at the lexical level. As long as the retrieved vector and the primed item are positively correlated in any case[7] activating or priming the lexical item (which I take to mean increasing the item's power or self dot product) will increase the magnitude of its resonance to the retrieved item. This item will now be chosen in preference to other possibilities that might have been chosen had the correct alternative not been primed. Thus, recall may be enhanced by priming that has no effect on the episodic composite trace.

Insofar as the feeling-of-knowing judgment is based on the quick monitoring of the familiarity between the cue and the composite trace, target priming need not affect these judgments. Of course, if feeling-of-knowing judgments were based exclusively on the partial retrieval of target information, then anything that enhances that retrieval should also increase the feeling of knowing. By this second hypothesis, Jameson et al.'s (1990) finding that target priming enhanced recall but had no effect on feeling-of-knowing ratings is inexplicable. The target priming results, then, provide strong support for a preretrieval locus of the feeling-of-knowing judgments.

Knowing Not

Kolers and Palef (1976) made the simple but provocative observation that people know what they do not know. Lest that statement seem paradoxical, let me elaborate. Suppose one were asked a question such as "What is the name of the largest department store in Budapest?" Many people would answer that they do not know; they would give a very low feeling-of-knowing rating. But it does not appear that they do so by trying to retrieve what they do know about large department stores in Budapest. Rather, they seem to have positive knowledge that they know nothing whatsoever about the topic. Empirically, these kinds of don't-know judgments are made very quickly.

This result is difficult to interpret under the assumption that people first retrieve and then piece together the don't-know judgment from partial information. It is more plausible to assume that the value from a preretrieval familiarity monitor, such as is given in the model, is used. Kolers and Palef (1976) criticized the idea that negative judgments could be based on a familiarity mechanism. They noted

> One way to handle knowing not is in terms of familiarity, in an extrapolation of the proposal of Atkinson et al. (1974). In that extrapolation one would recognize rapidly that one did not know an item, as one would recognize rapidly that one did know an item, and items of intermediate familiarity would be recognized more slowly. This seemingly plausible proposal is logically faulty, however; it supposes that a person would have in mind, for purposes of matching or checking, all of the items that he does not know. This is self-contradictory. (p. 557)

The argument of Kolers and Palef seems to be based on the assumption of specific retrieval. If retrieval (i.e., bringing again to consciousness the memorial token of the event) were necessary to make a don't-know judgment, then an explicit representation of the unknown information would be necessary, and the contradiction to which Kolers and Palef refer would be implied. However, with a familiarity monitor of the sort outlined earlier, the value the monitor returns is low for new and unrelated information and high for old or related information, and no specific representation of the unknown information is required to already exist in the system. With this kind of familiarity monitor, Kolers and Palef's criticism does not apply. If the value returned by the monitor was extremely low, it would indicate the utter novelty of the question and would produce a quick don't-know judgment.

[7] If the correct lexical item is uncorrelated or negatively correlated with the retrieved item, no amount of priming will produce a positive correlation. However, given that the effect starts out being positive, priming will enhance it.

Explicit Don't-Know Information

An experiment by Glucksberg and McCloskey (1981) provided an interesting follow-up to Kolers and Palef's (1976) work. Instead of relying only on implicit don't-know information, in one condition they specifically gave subjects statements indicating that certain information was not known. In the explicit conditions they gave subjects statements like "It is not known whether Gabriel owns a violin." If subjects were basing their judgments on what they retrieved, this explicit information should help; they should be able to say even more quickly that this information was not known.

On the other hand, consider the effect of explicit don't-know information if judgments are made on the basis of the cue-familiarity monitor. Presumably, an event quite similar to the question later used to cue memory ("Does Gabriel own a violin?") would have been entered into the trace. If the feeling-of-knowing judgment was based on the familiarity of the cue to the trace, the cue would now match to some extent. Critically, it would match more than if the explicit don't-know information had not been given. The familiarity monitor would produce a higher value, indicating that subjects should be less able to say that they did not know, in the explicit than in the implicit condition. To make a correct don't-know judgment in the explicit cases, subjects might even have to use a retrieval process rather than only using the familiarity monitor.

Glucksberg and McCloskey (1981) found that the explicit don't-know information slowed people down in making their don't-know judgments. Although they did not interpret the results in this manner, their findings are consistent with the conjecture that fast don't-know judgments can be made on the basis of the familiarity monitor. These slow judgments when explicit information is given require retrieval, however, and this explicit retrieval is a slower process. (The similarity of this suggestion to the two-stage models of recognition proposed by Atkinson & Juola, 1973, and by Mandler, 1980, is noteworthy.)

Summary

These eight lines of evidence converge on the conclusion that feeling-of-knowing judgments may not be based only on retrieved partial information. Instead, such judgments may be based on a monitoring mechanism that quickly assesses the familiarity–novelty of the event. In the model, such a monitoring mechanism is necessary as a precursor to memory storage to keep the variance of the memory trace within reasonable bounds. If feeling-of-knowing judgments were based on a monitoring-control mechanism, then under some conditions, one might expect to find that this mechanism is impaired. The next section of this article addresses this possibility and some implications not only for feeling-of-knowing judgments per se, but also for other aspects of human memory performance.

The Link to Korsakoff Amnesia

Feeling-of-knowing judgments are robust under a wide variety of experimental and subject variables, as mentioned earlier. Furthermore, these judgments are not necessarily linked to the absolute level of recall or recognition. Even though subjects may be capable of remembering only a few items, they can predict in advance which items these will be. However, as described later, there is one patient population unable to give reliable feeling-of-knowing judgments. This finding provides additional converging evidence that this particular metacognitive process may have a specific locus, separable from the basic memory retrieval system.

Feeling-of-Knowing Judgments

Shimamura and Squire (1986) tested the hypothesis that Korsakoff patients are selectively impaired in making feeling-of-knowing judgments. These patients seem to differ from other amnesic patients and from normal persons along a number of possibly related dimensions. First, as Shallice and Evans (1978) have pointed out, Korsakoffs have difficulty making estimates about everyday objects or events (e.g., "How tall is the average English woman?"). They tend to be apathetic and to show poor performance on categorization tasks such as the Wisconsin Card Sorting Test. This deficit is often taken as a clinical marker of frontal lobe dysfunction. Two major brain regions are especially implicated in Korsakoff amnesia—the frontal lobes and the diencephalon (Jackobson & Lishman, 1987; Shimamura, Jernigan, & Squire, 1988). Usually monitoring, planning, and metacognitive functioning are considered to involve the frontal lobes critically.

Shimamura and Squire's (1986) first experiment compared Korsakoff amnesic patients with other amnesic patients: electroconvulsive therapy (ECT) patients and 4 patients with organic amnesia not attributable to Wernicke-Korsakoff's syndrome. One of the non-Korsakoff amnesics was N.A., who suffers from diencephalic damage without the frontal complications seen in Korsakoff amnesia. General information questions such as "Who painted 'Afternoon at La Grand Jatte'?" (Seurat) were used in the first experiment. Korsakoff patients were impaired somewhat more than were the other amnesics and the alcoholic control subjects on the recall part of the task. Recognition was about the same for all groups, however. The major finding of interest was that the Korsakoff patients were severely impaired on the feeling-of-knowing task, and the other patients performed at the normal level on the metacognitive task. Data from this experiment are presented in Figure 7.

In Shimamura and Squire's (1986) second experiment, rather than using general information questions that may have been learned before the onset of the disease (in the case of the Korsakoff patients) or the trauma (in the case of the non-Korsakoff amnesics), sentences were given to subjects in the experimental setting (e.g., "At the museum we saw some relics made of clay."). Both the Korsakoff patients and the other amnesics suffered a similar, and severe, degree of amnesia on the immediate test. The alcoholic patients and the control subjects, though showing very good performance on immediate testing, were equated with the amnesic patients on memory performance by interposing a delay interval of between 1 and 7 days between the study and test. The non-Korsakoff amnesia patients as well as the normal patients who were tested at a delay were well above chance on their feeling-of-knowing correlations to performance. The Korsakoff patients, however, produced a gamma

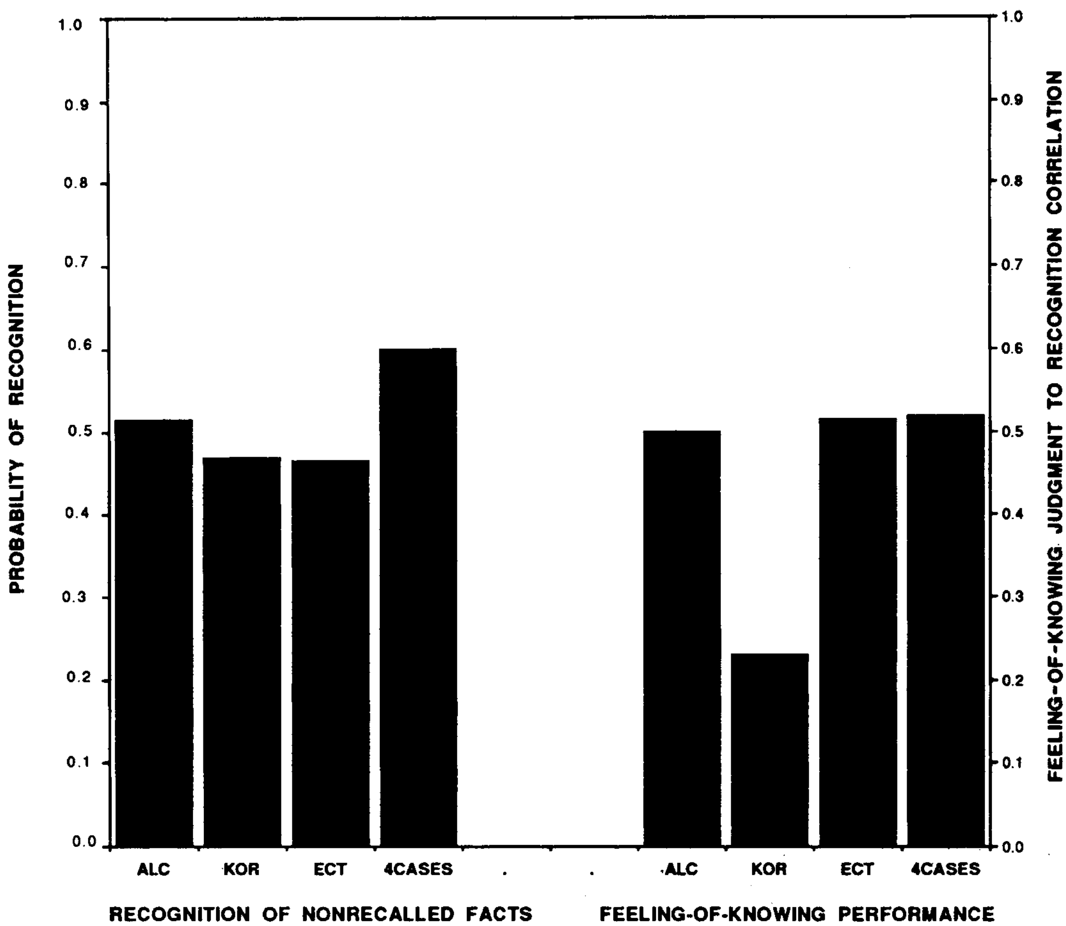

Figure 7. Feeling-of-knowing data for Korsakoff and non-Korsakoff patients. (ALC = alcoholic patients; KOR = Korsakoff amnesics; ECT = electroconvulsive therapy; 4 = 4 non-Korsakoff amnesics. Adapted from "Memory and Metamemory: A Study of the Feeling-of-Knowing Phenomenon in Amnesic Patients," by A. P. Shimamura and L. R. Squire, 1986, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* p. 455. Copyright 1986 by the American Psychological Association. Adapted by permission.)

correlation of −.007—showing no discernible ability to make accurate judgments about what they would be able to remember.

N.A. showed no feeling-of-knowing impairment and had diencephalic damage but no frontal damage. Korsakoff amnesics tended to show both diencephalic and frontal lobe damage and showed impairment on the feeling-of-knowing task. Taken together, these results suggest that frontal lobe damage was critical to the impairment in feeling-of-knowing judgments. In a follow-up study, Janowsky, Shimamura, and Squire (1989) tested a small group of patients whose only deficit was frontal lobe damage. The extent and the site of the lesions varied greatly over this small group. These patients experienced little memory impairment and therefore provide a contrast to the Korsakoff patients who had both metacognitive and memory impairments. Like the Korsakoff patients, these patients manifested selective impairment in their feeling-of-knowing judgments. The results with this small group were not as dramatic as those of Korsakoff patients. The metacognitive deficit was pronounced only at a 1- to 3-day delay of testing. As Janowsky et al. (1989) pointed out,

> Due to the variability in the locus and extent of damage in this small group of patients, it was not possible to explore statistically how feeling-of-knowing accuracy was affected by site and size of lesions. It is worth pointing out, however, that the 2 patients with bilateral frontal lesions, who achieved 0 and 1 categories on the Wisconsin Card Sorting Test, also achieved very low scores on the feeling-of-knowing test. . . . The findings from Experiment 1 suggest that the ability to make accurate feeling-of-knowing judgments depends, at least in part, on the integrity of the frontal lobes. (p. 7)

The results from Korsakoff and from frontal lobe patients indicate that the monitoring function that, in the model, keeps track of the variance of the trace, may under some circumstances be selectively impaired. This selective impairment certainly suggests that this function, which is a separable system from the basic memory model, is also separable in humans. In the next section, I investigate the results for recall either of implementing the model with an intact monitoring and control system or of disengaging this control system. Although Korsakoff and frontal lobe patients almost certainly retain partial monitoring ability, in the simulations I take this impairment to the most extreme limit—running the model without the monitoring function at all.

Release From Proactive Inhibition

One of the hallmarks of Korsakoff amnesics is that they fail to release from proactive inhibition (Cermak, Butters, & Morienes, 1974; Kinsbourne & Wood, 1975; Moscovitch, 1982; Squire, 1982; Warrington, 1982; Winocur, 1982; Winocur, Kinsbourne, & Moscovitch, 1981). In this task, patients are presented with a number of lists of items, each of which consists of items from the same category. After several such lists either there is a shift to a new category or there is no shift. When this task is conducted with normal subjects each list is usually followed by a counting backwards distracter task (following J. Brown, 1958; Peterson & Peterson, 1959; and Wickens, 1972). In studies with amnesics, the distracter task is often omitted because of the extreme difficulty amnesics experience with distraction. Winocur's (1982) data from such a release from proactive inhibition task, for Korsakoffs and a small group of normal control subjects are shown in Figure 8. Trial 5 was a shift trial, and as the dashed line in the figure illustrates, performance of the normal subjects bounced back to the high level of Trial 1. However, the Korsakoff patients showed almost no release on this trial. Although a number of research groups have shown this failure-to-release phenomenon, the Squire (1982) study is especially informative.

Squire (1982) investigated the release from proactive inhibition phenomenon across a variety of patient populations contrasted to a normal control group. The patient classifications were (a) depressives, (b) alcoholics, (c) ECT patients, (d) N.A., and (e) Korsakoffs. All of the patients showed some decrement in memory performance. The depressed subjects are of especial interest because depression can sometimes be confused with organic amnesias (especially early Alzheimer's disease). These patients showed a decrement in overall performance compared with the control subjects, but they released from proactive inhibition. Alcoholic patients showed a like pattern. Patients who were receiving bilateral ECT, sometimes thought to have an impact on the temporal lobes and hippocampus, showed impaired memory performance overall, compared to normal subjects. Their performance was also worse than that of depressed patients not receiving ECT. However, they released normally from proactive inhibition. Finally, N.A. was tested in this paradigm. As mentioned earlier, N.A. is a patient who had sustained left diencephalic (left dorsal thalamus) damage. This happened during an accident with a miniature fencing foil. Damage in this region is also characteristic of Korsakoff patients, though the latter also usually show lesioning in the frontal lobes. As Figure 9 shows, N.A.'s memory performance was im-

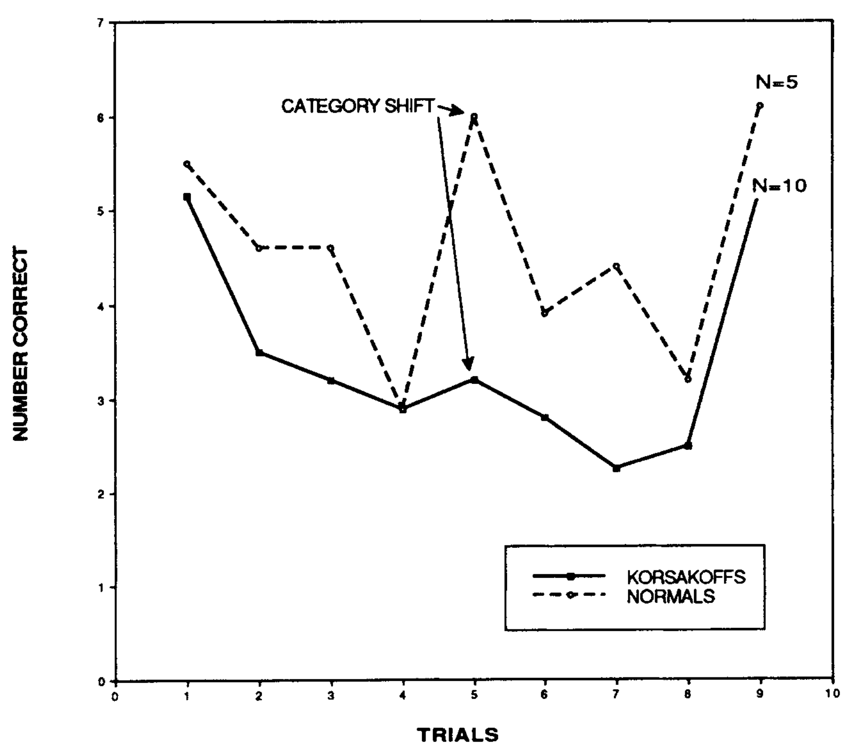

Figure 8. Release from proactive inhibition in normal subjects (broken line) and the failure to release in Korsakoff patients (solid line). (The category shift occurred on Trial 5. Adapted from "The Amnesic Syndrome: A Deficit in Cue Utilization," by G. Winocur, 1982, in L. S. Cermak, *Human memory and amnesia* [p. 150]. Hillsdale, NJ: Erlbaum. Copyright 1982 by Erlbaum. Adapted by permission.)

paired, but he released normally from proactive inhibition when given the category shift. In short, only the Korsakoff patients failed to release. These data are reproduced in Figure 9.

Moscovitch (1982) has compared a number of patient groups on the release from proactive inhibition task, paying special heed to frontal patients and to patients with hippocampal damage. One group that was not well represented in Squire's (1982) study was the temporal lobe and hippocampal group (although perhaps ECT patients show a deficit mainly due to impairment of function of this region). Moscovitch's data, shown in Figure 10, indicate that most patient groups (control subjects; temporal lobe patients, including those with pronounced hippocampal damage and severe memory deficits; and even right frontal lobe patients) released from proactive inhibition. Indeed, the only group showing a marked failure to release were patients with left frontal lobe damage. Interestingly, these patients also did extremely poorly on the Wisconsin Card Sorting Test, the same correlate that Janowsky et al. (1989) found to be related to poor performance on the feeling-of-knowing task.

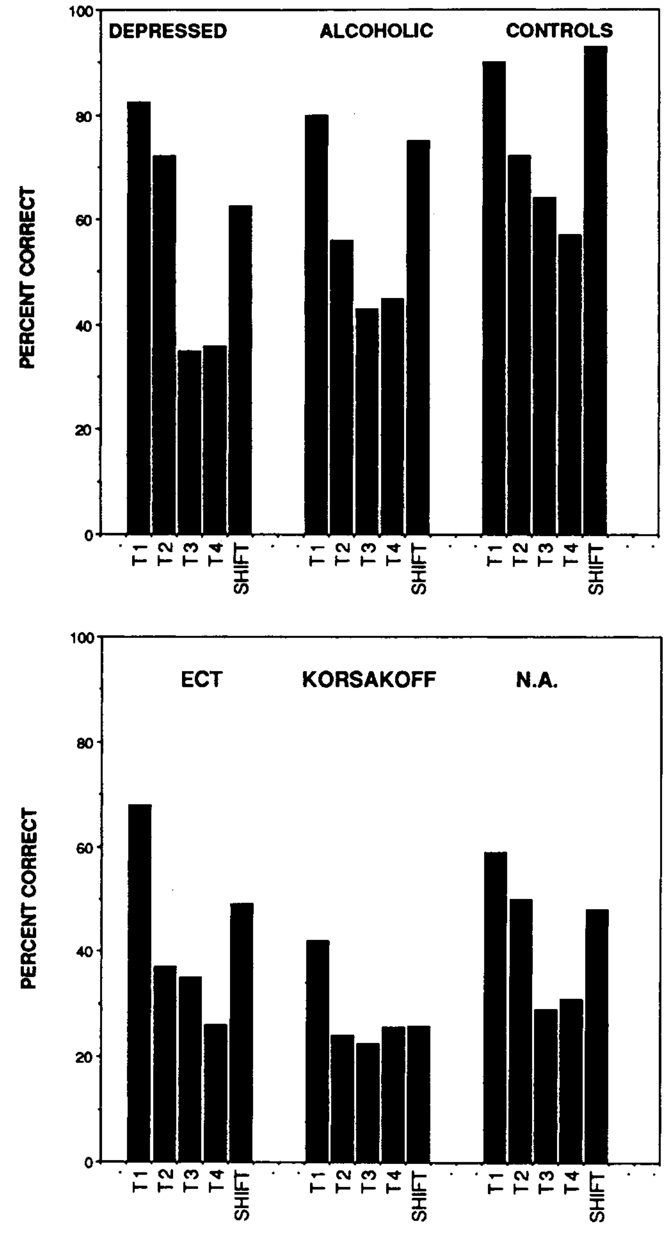

Figure 9. Release from proactive inhibition in various patient groups, including non-Korsakoff amnesics, and the failure to release in Korsakoff patients. (ECT = electroconvulsive therapy. Adapted from "Comparisons Between Forms of Amnesia: Some Deficits Are Unique to Korsakoff's Syndrome," by L. R. Squire, 1982, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* p. 564. Copyright 1982 by the American Psychological Association. Adapted by permission.)

Simulations of the Release From Proactive Inhibition Paradigm, With and Without Monitoring and Control

Until now I have not allowed the model to recall. In this final simulation, the recall performance of the model is compared with and without the control on the trace imposed by the novelty–familiarity monitor and filter. Categorized lists were constructed in a manner like that outlined in the Novelty–Familiarity Monitoring and Filtering section concerning trace renormalization as illustrated in Figure 6. In all cases, the composite trace started with a noisy vector, rather than a zero vector, to allude to the fact that there is always something in memory at the start of any experiment; it would be inappropriate to ever imagine that memory is a blank slate. After the presentation of each new (though categorically related) pair, the model was provided with the cue member of the last presented pair, which was correlated with the trace. The vector that was produced by this retrieval scheme was then matched to each of the items in the lexicon (consisting of 70 items, including four categories each of 10 items, and 30 additional unrelated items). The lexical item that exhibited the highest resonance score with the retrieved item was the item that was said to be recalled. If it was the target item, then recall was said to have been correct.

There were two manipulations of interest in the simulations. First, in half of the simulations the variance of the trace was computed with each new entry and the trace was renormalized. This is called the *monitored–controlled* condition. In the other half of the simulations no monitoring or control device was engaged. This is the *unmonitored–uncontrolled* condition. The second manipulation was crossed with the first. The second manipulation was the number of features in the vectors used in the model. A number of previously published investigations of the model have shown that with increasing number of features performance in the model improves (e.g., Metcalfe, 1991b). With this simple manipulation, then, diffuse basic memory system (i.e., hippocampal) amnesia can be mimicked (see Hirst, 1982; Squire, 1987; Weiskrantz, 1987) because with fewer features overall performance will be impaired. In half of the simulations, 63 features were used in the lexical representations and in the memory vectors, whereas in the other half, only 31 features were used.

As the top panel of Figure 11 illustrates, the model released from proactive inhibition in a normal manner when the monitoring–control device was in place. The variation in the number

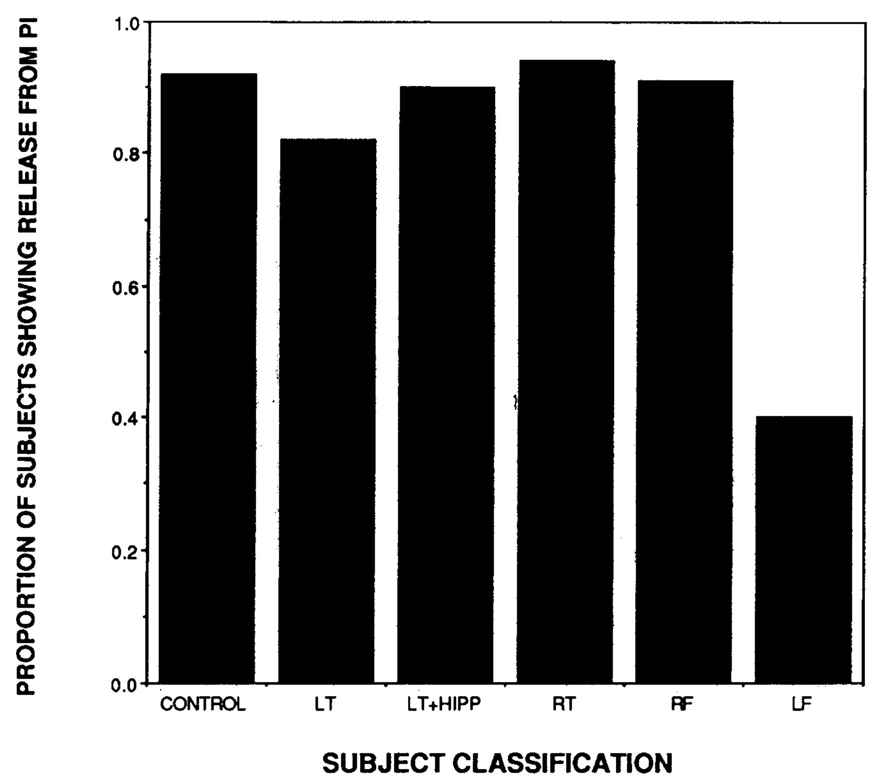

Figure 10. Proportion of subjects who showed release from proactive inhibition (PI). (The groups are as follows: control subjects, left temporal lobe patients [LT], left temporal lobe including hippocampal lesion patients [LT + HIPP], right temporal lobe patients [RT], right frontal lobe patients [RF], and left frontal lobe patients [LF]. Adapted from "Multiple Dissociations of Function in Amnesia," by M. Moscovitch, 1982, in L. S. Cermak, *Human memory and amnesia* [p. 345]. Hillsdale, NJ: Erlbaum. Copyright 1982 by Erlbaum. Adapted by permission.)

of features had an impact on the level of recall: With more features the overall levels were much higher than with fewer features. In both cases, however, the positive pattern of release was the same. These simulated results are reminiscent of those found by Squire (1982) with non-Korsakoff amnesics as compared with normal subjects. The bottom panel of Figure 11 shows the results without the monitoring–control device. Even with many features, overall performance was fairly low. Performance with few features, as would be expected, was even lower. However, the most striking feature of the results without a monitoring–control mechanism in place is that, given a category shift, the model failed to show release from proactive inhibition.

Discussion

The amnesia data presented here were already well established. For example, it was already known that frontal damage and Korsakoff syndrome are linked to impairments on the release from proactive inhibition task, among other tasks (see, e.g., Damasio, 1985; Eslinger & Damasio, 1985; Fuster, 1985; Goldman-Rakic, 1987; Levine & Prueitt, 1989; Milner, Petrides, & Smith, 1985; Schacter, 1989; Shallice, 1988; Shallice & Evans, 1978; Wilkins, Shallice, & McCarthy, 1987). One might ask, do researchers learn anything more from a very specific process model that can perform such a task (and others) and that gives rise to the same impairment effects as do patients? I think the answer is that modeling is essential to understanding how the mind–brain works—the agenda at the heart of both cognitive neuroscience and cognitive psychology. Localization findings, taken alone, tell one only that some test result was adversely (or, indeed, favorably) affected by injury to some brain region. Such findings can be highly reliable but not provide understanding of the working of the system, its implications or interconnections. For example, the link between failure to release from proactive inhibition and the inability to make accurate feeling-of-knowing judgments is precarious at best, without some coherent theoretical mechanism (such as that proposed here) that connects these seemingly disparate findings.

To give an analogy, imagine applying the lesioning strategy to a car rather than a brain. Suppose one blowtorched certain areas and then ran a test battery, including such tasks as turning the key, trying the headlights, honking the horn, opening the door, rolling down the window, backing up, and so on. Some of these tasks might be of more significance than others, but one would not know which ones without a functional explanation of how the systems worked, interacted, and were tapped by the

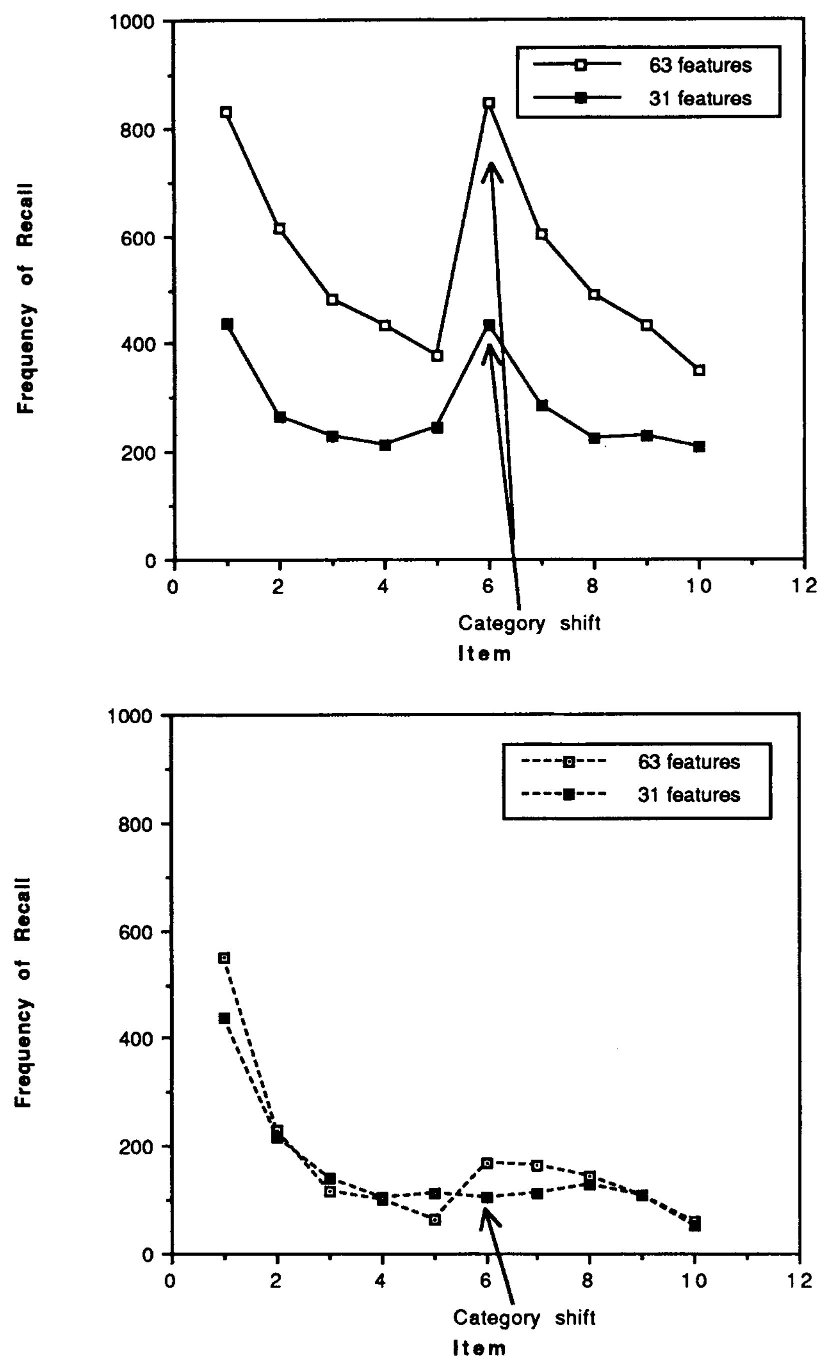

Figure 11. Simulation results in the release from proactive inhibition paradigm. (The top panel shows the model results with 63 or 31 features in the input vectors and with the monitoring device intact. In the bottom panel the monitoring-control mechanism is disengaged.)

tasks. One might also overlook the critical tests, but this likelihood is too depressing to consider further here. The blowtorch, of course, might easily hit more than one subsystem—the electrical system, for example, and the cooling system. However, if one had no theoretical understanding of what the systems were, one would be hard pressed to say whether one or multiple systems were responsible for the symptomology. In a particular case one might find the following: Given a blowtorch burn of some severity (quantified on the appropriate scale) 5 in. in back of the hood ornament and 14 in. to the left of center and, given that the car has been turned on for x min, smoke starts pouring out of the hood. The car runs, and the horn, lights, windows, and backup tests are all normal. The smoke effect appears to be reliable because it replicates on most but not all cars provided that the exact burn location is appropriately scaled for car size. Unfortunately, after having repeated the test results several times in one torched car, smoke fails to pour out, and a vile smell results, grinding is heard, and the car seizes up. Follow-up studies, using different cars, indicate that the initial hypothesis that it was smoke pouring out was incorrect—steam. Also, the sequential effect of the steam for a fixed interval followed by smell (which further experiments ascertain to be a joint function of the engine-'on' time and absolute time from the first indication of torching, a trade-off fitted by a mathematical function with a minimal number of parameters) appears not to be accidental but is also reliable. One concludes that the vector location (−14, 5) taken from the origin of the hood ornament must be the "steam, seize up, and sometimes smell generator."

Of course, if researchers know about the workings of the cooling system, then we might attain a deeper understanding. The radiator hose was torched, with the implications that follow from a mechanistic understanding of the function, mechanism, and locations of the cooling system and of its interaction with other systems. The symptoms follow, we know why they follow, and we can make further predictions. In short, the neuroscience-localization data, the psychological data, and the mechanistic theory are all needed to allow us to bootstrap our way to a better understanding of the human cognitive system. None of these alone is sufficient.

Conclusion

These investigations specified a mechanism that controls and modulates human memory and outlined some of the implications of this control-monitoring mechanism. One repercussion of such a monitoring mechanism is that it provides an informational basis that is quick and that does not depend on explicit retrieval for metacognitive judgments of feeling of knowing. Much of the data on these judgments is inexplicable under the alternate hypothesis that they are based on explicitly retrieved information. The investigations also indicate that not all deficits in memory are directly ascribable to the basic memory system. Failure in the adjunct control mechanism, conceptually distinct from the basic memory system, can produce amnesialike effects. Such effects, though, cluster into a syndrome that is distinguishable in detail from a syndrome resulting from impairments in the basic memory system. The model, then, provides a theoretical blueprint of the syndrome expected with control-based amnesia. The data presented here correspond to that blueprint, but further predictions of the model are left for exploration in future work with Korsakoff and frontally impaired patients.

References

Anderson, J. A. (1977). Neural models with cognitive implications. In D. LaBerge & S. Samuels (Eds.), *Basic processes in reading* (pp. 27–90). Hillsdale, NJ: Erlbaum.

Anderson, J. A., & Hinton, G. E. (1981). Models of information processing in the brain. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 9–48). Hillsdale, NJ: Erlbaum.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413–451.

Atkinson, R. L., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583–612). San Diego, CA: Academic Press.

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior, 12,* 311–319.

Brown, A. L., & Lawton, S. C. (1977). The feeling of knowing experience in educable retarded children. *Developmental Psychology, 13,* 364–370.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12–21.

Cavanagh, P. (1976). Holographic and trace strength models of rehearsal effects in the item recognition test. *Memory & Cognition, 4,* 186–199.

Cermak, L. S., Butters, N., & Morienes, J. (1974). Some analyses of the verbal encoding deficits of alcoholic Korsakoff patients. *Brain and Language, 1,* 141–150.

Cottrell, G. W., & Metcalfe, J. (1991). EMPATH: Face, emotion, and gender recognition using holons. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems* (Vol. 3, pp. 564–571). San Mateo, CA: Kaufmann.

Cultice, J. C., Somerville, S. C., & Wellman, H. M. (1983). Preschoolers' memory monitoring: Feeling of knowing judgments. *Child Development, 54,* 1480–1486.

Damasio, A. R. (1985). The frontal lobes. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (2nd ed.). New York: Oxford University Press.

Darley, C. F., Tinklenberg, J. R., Roth, W. T., Vernon, S., & Kopelt, B. S. (1977). Marijuana effects on long-term memory assessment and retrieval. *Psychopharmacology, 52,* 239–241.

Dennett, D. C. (1985). *Elbow room.* Cambridge, MA: MIT Press.

Eslinger, P. J., & Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology, 35,* 1731–1741.

Farah, M. J. (1990). *Visual agnosia.* Cambridge, MA: MIT Press.

Feller, W. (1957). *An introduction to probability theory and its applications.* New York: Wiley.

Freedman, J. L., & Landauer, T. K. (1966). Retrieval of long-term memory: "Tip of the tongue" phenomenon. *Psychonomic Science, 4,* 309–310.

Fuster, J. M. (1985). The prefrontal cortex: Mediator of cross-temporal contingencies. *Human Neurobiology, 4,* 169–179.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116,* 119–136.

Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 311–325.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational knowledge. In F. Plum & V. Montcastle (Eds.), *Higher cortical function: Handbook of physiology* (pp. 373–417). Washington, DC: American Physiological Society.

Gruneberg, M. M., & Monks, J. (1974). Feeling of knowing and cued recall. *Acta Psychologica, 38,* 257–265.

Hart, J. T. (1965). Memory and the feeling of knowing experience. *Journal of Educational Psychology, 56,* 208–216.

Heath, R. A., & Fulham, R. (1988). An adaptive filter model for recognition memory. *British Journal of Mathematical and Statistical Psychology, 41,* 119–144.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Hinton, G. E., & Anderson, J. A. (1981). *Parallel models of associative memory.* Hillsdale, NJ: Erlbaum.

Hirst, W. (1982). The amnesic syndrome: Descriptions and explanations. *Psychological Bulletin, 91,* 435–460.

Hockley, W. E., & Murdock, B. B. (1988). A decision model for accuracy and response latency in recognition memory. *Psychological Review, 94,* 314–358.

Jackobson, R. R., & Lishman, W. A. (1987). Selective memory loss and global intellectual deficits in Korsakoff's syndrome. *Psychological Medicine, 17,* 649–655.

Jameson, K. A., Narens, L., Goldfarb, K., & Nelson, T. O. (1990). The influence of near-threshold priming on metamemory and recall. *Acta Psychologica, 73,* 55–68.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology, 17,* 3–11.

Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General, 117,* 390–394.

Johnson, M. K. (1991, July). *MEM and source monitoring.* Paper presented at the International Conference on Memory, Lancaster, United Kingdom.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kinsbourne, M., & Wood, F. (1975). Short-term memory processes and the amnesic syndrome. In D. Deutsch & A. J. Deutsch (Eds.), *Short-term memory* (pp. 258–291). San Diego, CA: Academic Press.

Kohonen, T. (1977). *Associative memory: A system-theoretical approach.* Berlin, Federal Republic of Germany: Springer-Verlag.

Kohonen, T., Oja, E., & Lehtio, P. (1981). Storage and processing of information in associative memory systems. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 105–144). Hillsdale, NJ: Erlbaum.

Kolers, P. A., & Palef, S. R. (1976). Knowing not. *Memory & Cognition, 4,* 553–558.

Koriat, A. (1991, July). *A model of metamemory.* Paper presented at the International Conference on Memory, Lancaster, United Kingdom.

Krinsky, R., & Nelson, T. O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica, 58,* 141–158.

Lachman, J. L., Lachman, R., & Thronesbury, C. (1979). Metamemory through the adult life span. *Developmental Psychology, 15,* 543–551.

Lashley, K. S. (1950). In search of the engram. In *Symposia for the Society of Experimental Biology* (Vol. 4). Cambridge, England: Cambridge University Press.

Levine, D. S., & Prueitt, P. S. (1989). Modeling some effects of frontal lobe damage: Novelty and perseveration. *Neural Networks, 2,* 103–116.

Lewandowsky, S., & Murdock, B. B. (1989a). A distributed memory model for associative learning. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms, and models* (pp. 381–394). Amsterdam: North-Holland.

Lewandowsky, S., & Murdock, B. B. (1989b). Memory for serial order. *Psychological Review, 96,* 25–57.

Liepa, P. (1977). *Models of content addressable distributed associative memory (CADAM).* Unpublished manuscript. University of Toronto, Toronto, Ontario, Canada.

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

Marcel, A. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15,* 197–237.

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing* (Vol. 1). Cambridge, MA: MIT Press.

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 288–294.

Metcalfe, J. (1990). Composite holographic associative recall model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology: General, 119,* 145–160.

Metcalfe, J. (1991a). Composite memories. In W. Hockley & S. Lewandowsky (Eds.), *Relating theory and data* (pp. 399–423). Hillsdale, NJ: Erlbaum.

Metcalfe, J. (1991b). Recognition failure and the composite memory trace in CHARM. *Psychological Review, 98,* 529–553.

Metcalfe, J., & Bjork, R. A. (1991). Composite models never (well, hardly ever) compromise: Reply to Schooler and Tanaka (1991). *Journal of Experimental Psychology: General, 120,* 203–210.

Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 20,* 161–189.

Metcalfe, J., & Weibe, D. (1987). Insight and noninsight problem solving. *Memory & Cognition, 15,* 238–246.

Metcalfe Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Metcalfe Eich, J. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review, 92,* 1–38.

Milner, B., Petrides, M., & Smith, M. L. (1985). Frontal lobes and the temporal organization of memory. *Human Neurobiology, 4,* 137–142.

Moscovitch, M. (1982). Multiple dissociations of function in amnesia. In L. S. Cermak (Ed.), *Human memory and amnesia* (pp. 337–370). Hillsdale, NJ: Erlbaum.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B. (1985). Convolution and matrix systems: A reply to Pike. *Psychological Review, 92,* 130–132.

Nelson, T. O. (1988). Predictive accuracy of the feeling of knowing across different criterion tasks and across different subject populations and individuals. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 190–196). New York: Wiley.

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling of knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113,* 282–300.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676–686.

Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. S., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 279–288.

Nelson, T. O., McSpadden, M., Fromme, K., & Marlatt, G. A. (1986). Effects of alcohol intoxication on metamemory and on retrieval from long-term memory. *Journal of Experimental Psychology: General, 115,* 247–254.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26,* 125–171.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review, 91,* 281–294.

Reder, L. M. (1987). Selection strategies in question answering. *Cognitive Psychology, 19,* 90–138.

Reder, L. M. (1988). Strategic control of retrieval strategies. *The Psychology of Learning and Motivation, 22,* 227–259.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing* (Vol. 2). Cambridge, MA: Bradford.

Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 39–54.

Schacter, D. L. (1989). Memory, amnesia, and frontal lobe dysfunction. *Psychobiology, 15,* 21–36.

Schank, R. C. (1982). *Dynamic memory: A theory of reminding and learning in computers and people.* Cambridge, England: Cambridge University Press.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, England: Cambridge University Press.

Shallice, T., & Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex, 14,* 294–303.

Shimamura, A. P., Jernigan, T. L., & Squire, L. R. (1988). Radiological (CT) findings in patients with Korsakoff's syndrome and their relationship to memory impairment. *Journal of Neuroscience, 8,* 4400–4410.

Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: A study of the feeling of knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 452–460.

Squire, L. R. (1982). Comparisons between forms of amnesia: Some deficits are unique to Korsakoff's syndrome. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 560–571.

Squire, L. R. (1987). *Memory and brain.* New York: Oxford University Press.

Uttal, W. R. (1975). *An autocorrelation theory of form detection.* Hillsdale, NJ: Erlbaum.

Uttal, W. R. (1988). *On seeing forms.* Hillsdale, NJ: Erlbaum.

Warrington, E. K. (1982). The double dissociation of short- and long-term memory deficits. In L. S. Cermak (Ed.), *Human memory and amnesia* (pp. 61–76). Hillsdale, NJ: Erlbaum.

Weber, E. U. (1988). Expectation and variance of item resemblance distributions in a convolution-correlation model of distributed memory. *Journal of Mathematical Psychology, 32,* 1–43.

Weiskrantz, L. (1987). Neuroanatomy of memory and amnesia: A case for multiple memory systems. *Human Neurobiology, 6,* 93–106.

Wellman, H. M. (1977). Tip of the tongue and feeling of knowing experiences: A developmental study of memory monitoring. *Child Development, 48,* 13–21.

Wickens, D. D. (1972). Characteristics of word encoding. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 191–215). New York: Holt, Rinehart & Winston.

Wilkins, A. J., Shallice, T., & McCarthy, R. (1987). Frontal lesions and sustained attention. *Neuropsychologia, 25,* 359–365.

Winocur, G. (1982). The amnesic syndrome: A deficit in cue utilization. In L. S. Cermak (Ed.), *Human memory and amnesia* (pp. 139–166). Hillsdale, NJ: Erlbaum.

Winocur, G., Kinsbourne, M., & Moscovitch, M. (1981). The effect of cuing on release from proactive interference in Korsakoff amnesic patients. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 56–65.

Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 187–205.

Received September 28, 1990
Revision received October 10, 1991
Accepted March 13, 1992 ■