Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 2, 540–549

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.2.540

OBSERVATIONS

Turning up the Noise or Turning Down the Volume? On the Nature of the Impairment of Episodic Recognition Memory by Midazolam

Kenneth J. Malmberg
Iowa State University of Science and Technology

Rene Zeelenberg and Richard M. Shiffrin
Indiana University Bloomington

E. Hirshman, J. Fisher, T. Henthorn, J. Arndt, and A. Passannante (2002) found that Midazolam disrupts the mirror-patterned word-frequency effect for recognition memory by reversing the typical hit-rate advantage for low-frequency words. They noted that this result is consistent with dual-process accounts (e.g., R. C. Atkinson & J. F. Juola, 1974; G. Mandler, 1980; A. P. Yonelinas, 1994) of the word-frequency effect for recognition memory (S. Joordens & W. E. Hockley, 2000; L. M. Reder et al., 2000). The present authors show that this finding is also consistent with a variety of single-process, retrieving-effectively-from-memory (REM) models (R. M. Shiffrin & M. Steyvers, 1997), the simplest of which assumes that Midazolam decreases the accuracy with which memory traces are stored. These findings therefore do not discriminate between single- and dual-process models of recognition memory.

One way to study memory compares the performance of normal subjects to subjects with impairments (e.g., Cohen & Squire, 1980; Scoville & Milner, 1957; Warrington & Weiskrantz, 1970). Differences in performance can be used to infer that a given impairment might affect a specific brain mechanism or cognitive process that is thought to underlie memory. Memory theories based on such observations might, however, be constrained by the auxiliary assumptions concerning how a given impairment affects brain and cognitive processes. That is, deeper insights might be available by combining such empirical explorations with cognitive modeling. In this article, we use a computational-modeling approach to assess the necessity and sufficiency of certain auxiliary assumptions as the basis for explaining cross-population dissociations. Specifically, we use the retrieving-effectively-from-memory (REM; Shiffrin & Steyvers, 1997, 1998) modeling framework to address two issues. The first is a standard modeling question: How complex a model is needed to explain recognition memory data? The second is perhaps more interesting and more speculative: How do memory impairments degrade performance? That is, should memory impairments be characterized as conditions that reduce the amount of information stored in memory or should they be characterized as conditions that reduce the accuracy with which information is stored in memory?

A number of conditions lead to memory impairment (broadly defined), but the focus of this research is to explore ways for characterizing the impairments of *episodic recognition memory* due to the temporary influence of the drug Midazolam. (We further discuss this impairment below.) In an old–new recognition memory experiment, a list of items is studied, and participants are tested with both studied and unstudied items presented one at a time. The participant is asked to respond "old" to studied items and "new" to unstudied items. The probabilities of responding "old" to studied and unstudied items are referred to as the hit rate (HR) and the false-alarm rate (FAR), respectively. Hence, the impairment due to Midazolam is measured as a joint change to HRs and FARs relative to those obtained from a control participant (e.g., given saline).

Specifically, we consider how this impairment interacts with the effects of normative word frequency on recognition memory: Words differ in the frequency with which they are typically encountered in everyday life (Francis & Kučera, 1982). Relatively common and uncommon words are referred to as high- and low-frequency (i.e., HF and LF) words, respectively, and this factor has a robust effect on recognition memory: LF words are better recognized than HF words such that the HR is higher and FAR is lower for LF words than for HF words (Schulman, 1967; Shepard, 1967). This pattern of HRs and FARs is known as a "mirror effect" (Glanzer & Adams, 1985).

The empirical question investigated by Hirshman, Fisher, Henthorn, Arndt, and Passannante (2002) was how Midazolam affects the mirror-patterned word-frequency effect. They found that Midazolam selectively altered the pattern of HRs by reversing the typical LF advantage and had little or no effect on the LF-FAR advantage. Our specific objective is to explore single-process explanations for this pattern of disruption and to characterize the assumptions required for such an explanation.

According to single-process models of recognition memory (e.g., Dennis & Humphries, 2001; Hintzman, 1988; McClelland &

Kenneth J. Malmberg, Department of Psychology, Iowa State University of Science and Technology; Rene Zeelenberg and Richard M. Shiffrin, Department of Psychology, Indiana University Bloomington.

Correspondence concerning this article should be addressed to Kenneth J. Malmberg, Department of Psychology, Iowa State University, Ames, IA 50011. E-mail: malmberg@iastate.edu

Chappell, 1998; Murdock, 1997; Shiffrin & Steyvers, 1997, 1998), like REM, the old–new decision is based solely on the comparison of an item's familiarity to a criterion. If an item's familiarity exceeds an old–new criterion, an "old" response is made.

Hirshman et al. (2002) interpreted their findings within the framework of a dual-processing account of recognition memory (Atkinson & Juola, 1974; Mandler, 1980; Yonelinas, 1994). According to these accounts, old responses occur either when an item's familiarity exceeds a subjective criterion or when studying an item is explicitly recalled. According to some dual-process accounts of the word-frequency effect (Joordens & Hockley, 2000; Reder et al., 2000), the LF-HR advantage occurs because prior experiences with LF words are more likely to be recalled than prior experiences with HF words. This produces the LF-HR advantage. The LF-FAR advantage occurs because HF words are inherently more familiar than LF words, and therefore unstudied HF words are more likely to be called "old" in the absence of recalling the study event.

The findings of Hirshman et al. (2002) are therefore consistent with the hypothesis that Midazolam selectively disrupts recalling past experiences. Nevertheless, it is unclear from Hirshman et al.'s analyses whether their findings can distinguish between dual- and single-process models of recognition memory. One objective of the present research is to determine whether the dual-process model is necessary or merely sufficient for explaining the disruption of LF-HR advantage. Our approach is to discover whether and under what assumptions Shiffrin and Steyver's (1997, 1998) REM model might predict the interactions between normative word-frequency and Midazolam and aging.

As mentioned above, we entertain two hypotheses within REM's framework: Midazolam causes memories to be stored less accurately, or it causes less information to be stored in memory. The remainder of this observation is laid out in the following manner. We first take a closer look at the Hirshman et al. (2002) experiment and then describe a mechanism by which the simplest version of REM (Model 1.0 described by Shiffrin & Steyvers, 1997) can account for their findings. Finally, we briefly discuss more complicated versions of REM and implications that can be drawn from them.

Effect of Midazolam on Old–New and Remember–Know Recognition Memory

Study of the role of the medial temporal lobe (i.e., hippocampus, parahippocampus, perirhinal cortex, entorhinal cortex, and dentate gyrus) in human memory typically involves observing humans with lesions in this area often caused by aspiration, ablation, or chronic disease. Such studies have identified a critical role of the medial temporal lobe, because its damage leaves short-term memory function intact and leaves retrieval of previously learned material intact but prevents the storage of information in, or retrieval of information from, long-term memory (Squire, 1987). It would obviously be useful to have a technique to investigate these matters in humans with an undamaged medial temporal lobe. Hirshman and his colleagues (Hirshman et al., 2002; Hirshman, Passannante, & Henzler, 1999; also see Polster, McCarthy, O'Sullivan, Gray, & Park, 1993) recently developed one such technique: Participants are administered Midazolam, a benzodiazepine that temporarily causes anterograde amnesia, with effects that generally mimic those found after medial temporal lobe damage.

The participants in the experiment conducted by Hirshman et al. (2002) studied lists of HF and LF words, and the amount of time allocated for study was varied (either not studied or studied for 500, 1,200, or 2,500 ms per word). In addition, participants received either saline or Midazolam prior to studying the list of words. After a delay of about an hour, they were given an old–new recognition test and a remember–know recognition test (which we describe below).

The HRs and FARs from Hirshman et al.'s (2002) experiment are depicted in Figure 1. The points labeled with zero study time give FARs, and the other points give HRs. The results from the saline condition, given in the upper panel, replicate the standard effects in the literature. Better performance was observed for LF words: For LF words, FARs were lower and HRs were higher, and performance improved with increases in study time. The Midazolam group, of course, gave lower performance, but the pattern of results also differs from that for the saline group. Although LF performance was (slightly) better than HF performance, the mirror effect was lost: LF words gave lower FARs (as usual, and the size of this effect was similar to that for the saline group) but also lower HRs (not as usual). The effect of study time appeared diminished in the Midazolam condition, but the low levels of performance overall make comparisons problematic.

In addition to giving old–new recognition judgments, Hirshman et al. (2002) asked subjects to indicate whether their old judgments were made on the basis of "remembering" the study event or on the basis of "knowing" the word was studied even though they could not explicitly remember the study event (cf. Gardiner, 1988). The rationale for using remember–know judgments and interpretation of the results has been the subject of much debate (e.g., Donaldson, 1996; Donaldson, MacKenzie, & Underhill, 1996; Gardiner, 1988; Gardiner & Gregg, 1997; Hirshman & Henzler, 1998; Hirshman & Master, 1997). Tulving (1983) originally proposed that remembering and knowing reflect differences in the type of information retrieved from memory. Gardiner (1988) extended this idea to assess the viability of dual-process models of recognition memory (Atkinson & Juola, 1974; Mandler, 1980; Yonelinas, 1994). Such models typically propose that a familiarity process gives rise to a continuously distributed feeling of knowing without details of what has been retrieved and a recall process that returns details of a remembered event. Thus, participants are asked to rate their recognition responses to assess the contributions of these two components, and factors that differently affect the proportion of "remembering" and "knowing" responses have different effects on familiarity and recall processes.

The probabilities of responding either "remember" or "know" in Hirshman et al.'s (2002) experiment—conditional on making an "old" response—are shown in Figure 2. Of perhaps greatest interest for our present purposes, there was a large interaction between word frequency and drug manipulation: The probability of a "know" judgment was uniformly higher for HF than LF words, for both saline and Midazolam groups. However, the probability of a "remember" judgment was much higher for LF than HF words for

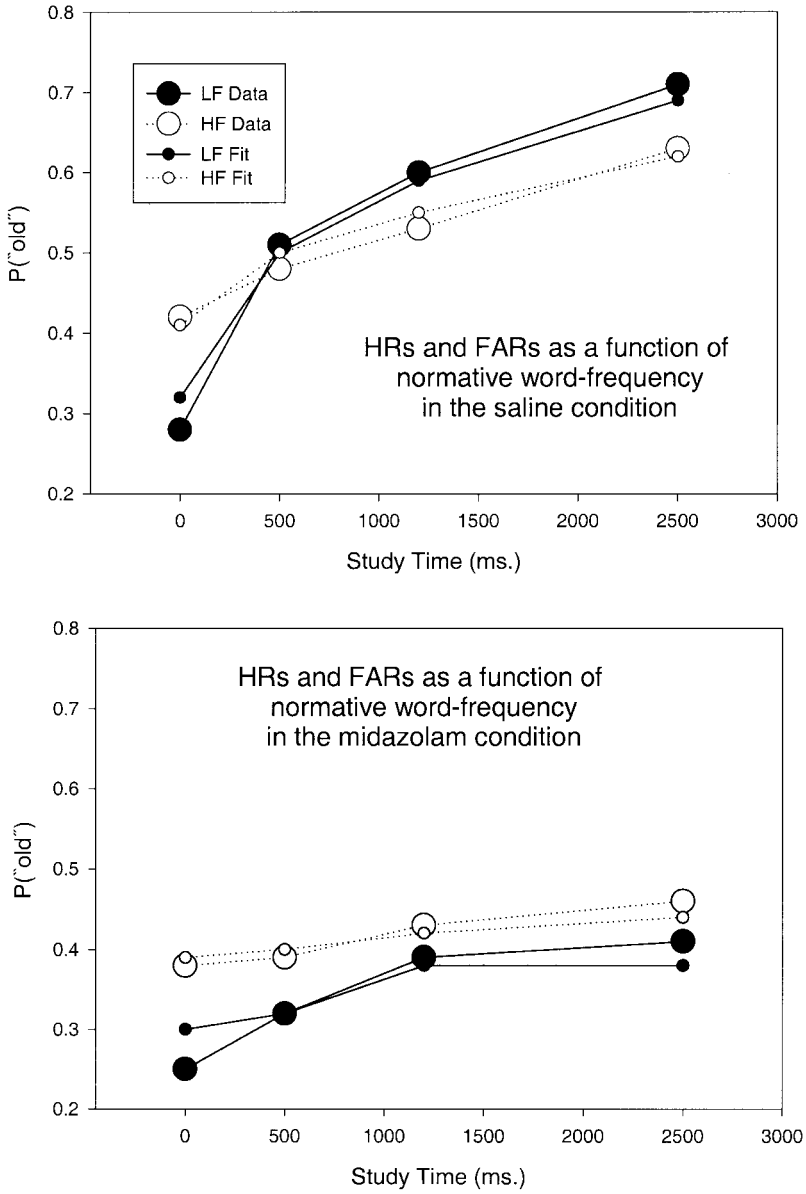

Figure 1. Yes–no recognition data from Hirshman et al.'s (2002) study and predictions of a retrieving-effectively-from-memory model. Zero-ms study time refers to "new" items so that the data give the false-alarm rate (FAR). Data shown for nonzero study times give hit rates (HR). Only the retrieving-effectively-from-memory parameter c varies between the saline and Midazolam conditions. The fits are based on 300 Monte Carlo simulations using $g_{LF} = .325$, $g = .40$, $g_{HF} = .45$, $w = 16$, $t_0 = 4$, $a = .8$, $u^* = .025$, $c_{\text{Sal}} = .77$, $c_{\text{Mid}} = .25$, and $\text{Crit}_{\text{O/N}} = .92$. LF = low-frequency words; HF = high-frequency words.

the saline group but did not differ for the Midazolam group (cf. Gardiner & Java, 1990; Joordens & Hockley, 2000; Reder et al., 2000).

Hirshman et al.'s (1999) new findings were that Midazolam causes the LF-remember advantage to disappear concomitant with the reversal of the LF-HR advantage. The results showing that LF words are more likely to receive "remember" responses than HF words are usually interpreted to be "consistent with" or to "support" the dual-process account of recognition memory (Gardiner & Java, 1990; Joordens & Hockley, 2000; Reder et al., 2000). Likewise, Hirshman et al. (2002) interpreted their new findings within a dual-process framework, suggesting—at least implicitly—that these results were inconsistent with single-process accounts such as those described by Donaldson (1996) and Hirshman and Master (1997). However, confirmation of the dual-process model does not necessarily mean disconfirmation of the single-process models, and in the following sections, we explore the single-process REM model in order to determine whether and under what conditions it

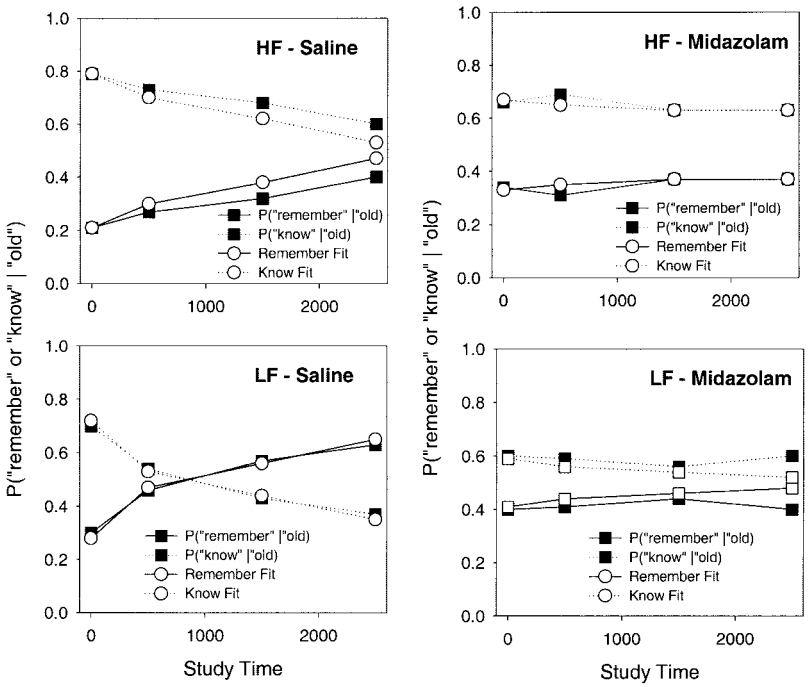

Figure 2. Remember–know data from Hirshman et al.'s (2002) study and predictions of a retrieving-effectively-from-memory model. Study time is in milliseconds. The fits are based on 300 Monte Carlo simulations using $g_{LF} = .325$, $g = .40$, $g_{HF} = .45$, $w = 16$, $t_0 = 4$, $a = .8$, $u^* = .025$, $c_{\text{Sal}} = .77$, $c_{\text{Mid}} = .25$, and $\text{Crit}_{\text{O/N}} = .92$. In addition, there are two remember-know criterion: For the saline group, $\text{Crit}_{\text{R/K}} = 1.52$; for the Midazolam group, $\text{Crit}_{\text{R/K}} = 1.30$. LF = low-frequency words; HF = high-frequency words.

might also predict the disruption of the LF-HR and the remembering advantages.

Single-Process Models of Old–New Recognition and Remember–Know Judgments

According to single-process models of recognition, performance is based on a continuous random variable that is often conceptualized as the strength, intensity, or familiarity associated with the test item (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Humphreys, Bain, & Pike, 1989; Kintsch, 1967; McClelland & Chappell, 1998; Murdock, 1993; Shiffrin & Steyvers, 1997). If the familiarity of the test item exceeds a subjective criterion, then the subject responds "old." Otherwise, a "new" response is made (cf. Green & Swets, 1966). A subclass of this type of model accounts for the word-frequency mirror effect by assuming that there exist four underlying distributions of familiarity values, such that the means of these distributions are arranged along a familiarity scale in the following manner: μ(LF-new) $<$ μ(HF-new) $<$ μ(HF-old) $<$ μ(LF-old). The left side of Figure 3 displays this relation graphically. A model of this type can predict the recognition findings of Hirshman et al. (2002) if the effect of Midazolam is to rearrange the underlying distributions on the familiarity scale such

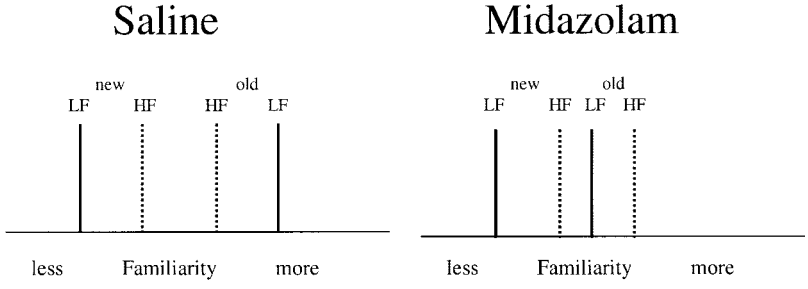

Figure 3. Arrangement of means of the theoretical distributions of strength-based models that may give rise to Hirshman et al.'s (2002) and Balota et al.'s (2002) findings. HF = high-frequency words; LF = low-frequency words.

that μ(LF-old) $<$ μ(HF-old). The right side of Figure 3 displays this relation graphically. The REM model of the word-frequency effect described by Shiffrin and Steyvers (1997, 1998; Malmberg, Steyvers, Stephens, & Shiffrin, 2002) is a member of this class of models, as we describe next.

REM

REM (Shiffrin & Steyvers, 1997) assumes that memory traces consist of vectors **V,** of length *w,* and of nonnegative integer feature values. Zero represents no information about a feature. Otherwise, the values for a given feature are assumed to follow a geometric probability distribution: $P(V = j) = (1 - g)^{j-1} g, j = 1,\ldots$. Thus, higher integer values represent less likely feature values. A vector produced with a lower value of g will tend to have higher values and hence tend to have features less likely to be encountered in the environment.

There are two types of memory traces in REM. Lexical–semantic traces represent general knowledge (e.g., the orthographic, phonological, semantic, and contextual characteristics of a word) and have very many nonzero feature values, most of which are encoded correctly. Episodic traces represent the occurrence of stimuli in a certain environmental context; they are built of the same feature types as lexical–semantic traces but tend to be incomplete (have many zero values) and inaccurate (the values do not necessarily represent correctly the values of the presented event).

REM assumes that HF words tend to consist of relatively common features and that LF words consist of relatively rare features, an assumption implemented by generating lexical–semantic feature values for HF words with one value of g (g_{HF}) and for LF words with a lower value of g (g_{LF}, where $g_{HF} > g_{LF}$). This assumption is known as the "feature-frequency" assumption (Malmberg et al., 2002).

When a word is studied, an incomplete and error prone representation of the word's lexical–semantic trace is stored in a separate episodic image. The probability that a feature will be stored in the episodic image after t time units of study is $1 - (1 - u^*)^t$, where u^* is the probability of storing a feature in an arbitrary unit of time. The number of attempts, t_j, at storing a content feature for an item studied for j units of time is computed from the following equation: $t_j = t_{j-1}(1 + e^{-aj})$, where a is a rate parameter and t_1 is the number of attempts at storing a feature in the first 1 s of study (Malmberg & Shiffrin, in press). Thus, increased study time increases the storage of features, but the gain in the amount of information stored diminishes as the item is studied longer. Features that are not copied from the lexical–semantic trace are represented by a value of 0.

If storage of a feature does occur, the feature value is correctly copied from the word's lexical–semantic trace with probability c. With probability $1 - c$, the value is incorrectly copied and sampled randomly from the long-run base-rate geometric distribution, a distribution defined by g such that $g_{HF} > g > g_{LF}$.

At test, a probe made with only context features is assumed to activate the episodic traces, I_j, of the n list items and no others (Shiffrin & Steyvers, 1997). Then, the content features of the probe cue are matched in parallel to the activated traces. For each episodic trace, I_j, the system notes the values of features of I_j that match the corresponding feature of the cue (n_{ijm} stands for the number of matching values in the jth image that have value i) and the number of mismatching features (n_{jq} stands for the number of mismatching values in the jth image). Next, a likelihood ratio, λ_j, is computed for each I_j:

$$\lambda_j = (1 - c)^{n_{jq}} \prod_{i=1}^{\infty} \left[\frac{c + (1 - c)g(1 - g)^{i-1}}{g(1 - g)^{i-1}} \right]^{n_{ijm}} \quad (1)$$

λ_j is the likelihood ratio for the jth image. It can be thought of as a match-strength between the retrieval cue and I_j. It gives the probability of the data (the matches and mismatches), given that the retrieval cue and the image represent the same word (in which case features are expected to match, except for errors in storage) divided by the probability of the data, given that the retrieval cue and the image represent different words (in which case, features match only by chance).

The recognition decision is based on the odds, Φ, giving the probability that the test item is old divided by the probability that the test item is new (Shiffrin & Steyvers, 1997). This is just the average of the likelihood ratios:

$$\Phi = \frac{1}{n} \sum_{j=1}^{n} \lambda_j \quad (2)$$

If the odds exceed a criterion, then an "old" response is made. The default criterion is 1.0 (which maximizes probability correct), although subjects could, of course, deviate from this setting.

Thus, an "old" response is given when there is more evidence that the test word is old. Matching features contribute evidence that an item is old (contribute factors to the product in Equation 1 greater than 1.0), and mismatching features contribute evidence that an item is new (contribute factors less than 1.0). REM predicts an effect of study time because storage of more nonzero features increases the number of matching target-trace features; this factor outweighs the general increase in variance produced by greater numbers of nonzero features in all vectors. REM predicts a LF-HR advantage because the matching of the more uncommon features associated with LF words produces greater evidence that the item is old than the matching of the more common features associated with HF words. For foils, however, every feature match is due to chance; such matching occurs more frequently for HF than LF words because HF features are more common (cf. Malmberg & Murnane, 2002). This factor outweighs the higher diagnosticity of matches for the LF words, and HF words are predicted to have higher FARs than LF words.

An REM Account for the Effect of Midazolam on Recognition

It is widely believed that the hippocampus region plays a critical role in storing episodic memory traces (e.g., Buzsaki, 1989; Marr, 1971; McClelland, McNaughton, & O'Reilly, 1995; O'Keefe & Nadel, 1978; Squire, 1987), and Midazolam has been shown to

affect the storage, but not the retrieval, of memory traces (Polster, et al., 1993). As described above, there are two parameters in REM that affect the storage of features in memory: u^* determines the number of features that get stored, and c determines the accuracy with which features get stored. To lower performance, it could be assumed that Midazolam reduces the values of either or both of these parameters. However, Hirshman et al.'s (2002) data constrain which of these possibilities is viable in the simplest form of the REM model of recognition memory (Shiffrin & Steyvers, 1997).

Let us assume that Midazolam only causes the hippocampal region to store fewer features, relative to the saline condition (i.e., u^* is reduced). In REM, this causes fewer terms in the product given by Equation 1 and a lower value for the result, on the average. Hence, if Midazolam causes fewer features to be stored, subjects should approach chance-level performance for both HF and LF words: LF(FAR) ~ HF(FAR) ~ LF(HR) ~ HF(HR). This prediction is shown in the right panel of Figure 4, which shows that the HRs and FARs for HF and LF words equal .50 when no nonzero features are stored in memory (i.e., $u^* = 0$) and the criterion is set to 1.0. Figure 4 shows that increasing the probability of storing features at study produces two mirror effects: Greater storage produces greater HRs and lower FARs, and the HRs are greater and FARs are lower for LF than for HF words. Hence, increasing the number of stored features causes changes in both HRs and FARs with the HF and LF advantages initially increasing and then leveling off once a relatively high-level performance is reached. However, Hirshman et al. (2002) found that only the LF-HR advantage was affected by Midazolam (see Figure 1). Specifically, the LF-HR advantage was reversed, whereas the LF-FAR advantage was unaffected, which is a pattern of data not predicted by the assumption that Midazolam produces less storage at study within this simple version of REM. Thus, the effect of Midazolam is not to reduce the number of features that get stored according to this simple REM model.

Next, let us assume that Midazolam causes the hippocampal region to store "noisier" episodic traces, as opposed to traces with fewer nonzero features. In REM, this is accomplished by decreasing the value of the c parameter, which is the probability that a feature will be copied correctly from a word's lexical–semantic trace to the episodic trace. The left panel of Figure 4 shows the predicted pattern of HRs and FARs as a function of the accuracy of storage during study (i.e., changes in c). It shows that decreasing c has virtually no effect on the FARs, because these FARs are based on chance matches. However, decreasing c causes the LF and HF old-item distributions (see Figure 3) to approach the LF and HF new-item distributions; when the decrease is large enough, this factor must cause the LF and HF old-item distributions to reverse position. The reversal occurs because the HF retrieval cues used to probe memory have more common features (on average) than the LF retrieval cues, a factor that comes to dominate when the true "signal" (matching features in the target trace) begins to disintegrate into noise (due to lowering of c). Hence, in REM terms, a point is reached at which μ(LF-old) drops below μ(HF-old), and this is exactly what is demanded by the findings of Hirshman et al. (2002, see Figure 1).

In addition, these assumptions predict the changes with study time. If u^* is not changed by Midazolam, the number of features stored rises with study time as in the saline group. However, lowering of c tends to cause random storage, so that the number of correctly stored features increases only slightly with increases in study time (i.e., t). This factor greatly diminishes the changes in performance with changes in study time, as observed. However, with low c the difference between LF and HF targets is due to chance matching of features that differ in diagnosticity, and hence the difference remains fairly constant across changes in study time, also as observed.

Figure 1 shows predictions of an REM model incorporating the assumption that only c varies between the saline and Midazolam groups, and only at storage. For retrieval (see Equation 1), the

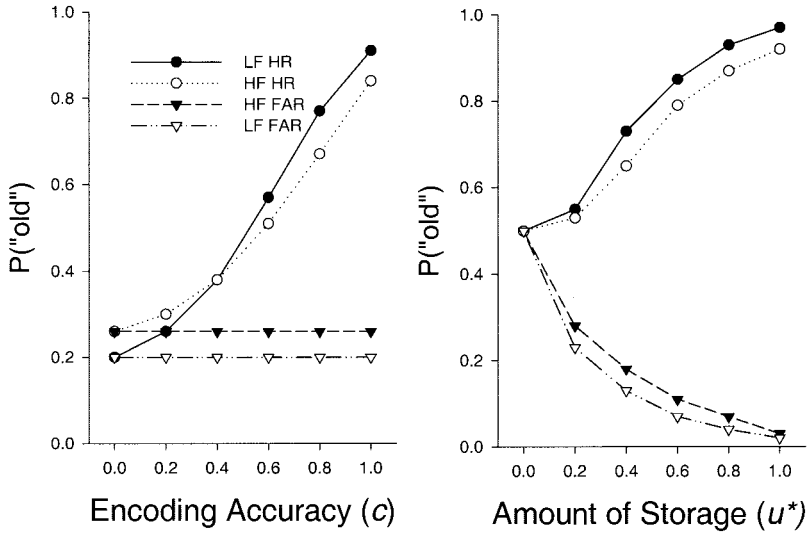

Figure 4. Retrieving-effectively-from-memory (Shiffrin & Steyvers, 1997) predictions for hit rates (HR) and false-alarm rates (FAR) as a function of the accuracy of storage (c) and the amount of storage (u^*). HF = high-frequency words; LF = low-frequency words.

same *c* value was used in both the saline and Midazolam conditions; this assumption seems conceptually consistent with the view that retrieval is tuned to the participant's learning over the course of development and accumulation of lifetime experience and is also consistent with prior findings showing that Midazolam affects the storage of traces and not their retrieval (Polster, et al., 1993). The criterion for an old–new judgment was set to .92, rather than the normatively optimal value of 1.0, in order to obtain a good quantitative fit, but the criterion did not vary between the Midazolam and saline groups and, therefore, is not of consequence for the present article. Clearly, the model provides a good fit of the data.[1] Hence, interpreted within the framework of REM, the main effect of Midazolam is to cause the hippocampal region to store more noisy episodic traces, implemented by a decrease in one parameter, *c*. These conclusions are based on the recognition data. We turn next to the remember–know judgments.

In accord with our stated goal, we chose to model remember–know judgments in what is probably the simplest way. The approach is based on the models described by Donaldson (1996) and Hirshman and Master (1997). As described above, an "old" decision is given when the familiarity (i.e., activation, or, in REM terms, the odds) associated with a test word exceeds the yes–no criterion. When this happens, then it is assumed that a higher remember–know criterion is set. Words whose familiarity exceeds the higher remember–know criterion are given the "remember" response, and a "know" response is given when the remember–know criterion is not exceeded. Figure 2 shows that this model predicts the effects of Midazolam and saline both qualitatively and quantitatively. This fit was obtained by using slightly different remember–know criteria in the saline and Midazolam conditions (1.52 and 1.30 in the saline and Midazolam conditions, respectively). It is important to note that all the qualitative effects are predicted correctly even when the same criterion is adopted for remember–know. These predictions provide an existence proof that single-process, familiarity-based models that use separate criteria for old–new and remember–know responses can account for both the reversal of the LF-HR advantage and the greater proportion of "remember" responses given to LF words than HF words.

Thus far, we have demonstrated the sufficiency of a model assuming that Midazolam reduces storage accuracy rather than storage quantity. What degree of mixture of these assumptions might be compatible with the data? An answer would require an exhaustive exploration of the parameter space, but we found that the use of a 50% reduced value of u^* for the Midazolam group ($u^*_{sal} = .02$; $u^*_{mid} = .01$) predicted an LF-FAR advantage that deviated from the data by being noticeably smaller in the Midazolam than saline condition. Within the REM framework, this result suggests that the main effect of Midazolam (possibly all the effect) is on *c* (accuracy of storage) rather than on u^* (quantity of storage).

General Discussion

The simple REM model we described predicts Hirshman et al.'s (2002) findings by assuming that Midazolam causes a decrease in accuracy of feature storage (rather than a decrease in number of features stored). Because Midazolam has a selective impairment in the hippocampus, we can speculate that Midazolam allows the hippocampal region to facilitate storage at a normal rate but that it disrupts the correctness of this storage.

There is at least one line of reasoning based on neuroscientific evidence that is consistent with this assumption: The hippocampus (proper) consists of approximately 10% GABAergic interneurons, and these interneurons are thought to control the firing of the remaining 90% of the hippocampal principle neurons (see Vizi & Kiss, 1998, for a review). Some of the principle neurons are granule neurons, and some are pyramidal neurons. The granule cells are associated with a rhythmic pattern of neuronal activity, known as theta waves (Buzsaki, 1989). Theta waves are associated with exploratory activities in both animals (O'Keefe & Nadel, 1978) and humans (Caplan, Madsen, Raghavachari, & Kahana, 2001), activities in which information about novel situations is being acquired. Midazolam is a benzodiazepine, and benzodiazepines inhibit the firing GABAergic interneurons in the hippocampus (Deadwyler, West, & Lynch, 1979). Hence, if Midazolam inhibits the firing of those cells that regulate the orderly firing of the vast majority of hippocampal cells, then it is reasonable to speculate that the result is a noisier episodic memory trace.

This speculation concerning the effect of Midazolam on the hippocampus raises questions about the effect of hippocampal lesions. Although such lesions could produce noisier storage, it seems superficially more plausible that lesions may instead cause fewer features to be stored (especially when the lesions are extensive). If so, then the effect of normative word frequency on lesioned subjects, or subjects with Korsakoff's syndrome, could be different than the effects reported by Hirshman et al. (2002). This question could be explored in future research.

The adequacy of the REM model to account for Hirshman et al.'s (2002) findings also raises the question of whether the impairment due to aging can be characterized as the encoding of less accurate episodic memory traces. For instance, Balota et al. (2002) recently reported the results of an experiment in which they varied the age of their subjects, the degree to which subjects were impaired by Alzheimer's disease (also see Wilson et al., 1983), and normative word frequency. The LF-HR advantage diminished with increases in age, and the LF-HR advantage was reversed in subjects suffering from mild Alzheimer's disease (i.e., the HR was greater for HF than for LF words), and LF-FAR advantage remained constant in both cases. Balota et al. interpreted their data as providing support for a dual-process account of recognition memory. On the assumption that recall- but not familiarity-based decisions become impaired with increases in age, these models predict a decrease or a reversal in the LF-HR advantage with increases in age as well as predicting that the LF-FAR advantage should remain intact, which is the pattern of data observed by Balota et al.

However, the prior sections show that even the simplest form of the REM model for recognition memory can predict Balota et al.'s

[1] Details of the modeling are reported in the figures. None of the reported fits are likely to be "best" fits of the model to the data. Rather, the fits show that the models could come reasonably close to the data (quantitatively as well as qualitatively). The larger symbols used to represent the data indicate that there is variability in the data, although the actual variability is not reported by Hirshman et al. (1999).

(2002) findings if it is assumed that the accuracy of storage decreases with increases in age.[2] Figure 4 illustrates this point qualitatively: Decreases in the REM *c* parameter produce decreases and, ultimately, a reversal in the LF-HR advantage and little or no change in the LF-FAR advantage. However, the details of the experimental design do not allow for meaningful quantitative model fits.[3]

Alternative REM Models of Recognition Memory

We have demonstrated that the simplest version of the REM model of recognition memory (Shiffrin & Steyvers, 1997) can predict the interaction between normative word frequency and Midazolam. According to this simple model, these factors cause noisier episodic traces to be stored. We also explored the possibility that Midazolam causes less storage in episodic memory, but such a model did not fit the data from Hirshman et al. (2002). However, Shiffrin and Steyvers (1997) described more complex (and, perhaps, more realistic) REM models that might not require the assumption that Midazolam causes storage to be less accurate but rather that it causes less storage. Here, we briefly discuss three such single-process models. The point of this discussion is not to determine which model is the "correct" model but rather to show that there are a variety of more complex ways that a single-process model can predict the findings of Hirshman et al.'s (2002) study.

The REM.5 model, described by Shiffrin and Steyvers (1997), is the same as the simplest REM model that we have discussed above except that it assumes that images consist of both item and context information and that memory consists of a large number of images that were stored prior to the experiment in addition to the images stored during study. We refer to the additional images as *extralist images.* According to the model, current context information is used to isolate an *activated set* of images in memory to which the retrieval cue is then compared; the more similar the context stored in an image, the more likely it will be a member of the activated set. Hence, extralist images are less likely to be a member of the activated set than images stored during study because those images stored during study are more likely to have context information that is similar to the test context.

The existence of extralist images in REM.5 opens up at least three new ways, in principle, to explain the Hirshman et al. (2002) findings. Each explanation is based on the following logic: Extralist traces will tend to be only randomly similar to the studied items, much in the same way that noisy images are only randomly similar to test items. Hence, one way REM.5 can, in principle, predict Hirshman et al.'s findings is to assume that Midazolam causes noisy images to be stored and that the activated set consists of a combination of extralist images and those images stored during study. However, a model that assumes extralist intrusions into the activated set can also make the proper prediction by assuming that Midazolam causes fewer features to be stored. According to this model, Midazolam leads to a reversal of the LF-HR advantage because HF words will tend to match a larger number of extralist traces than LF words. Lastly, Midazolam might cause a disruption in the binding of item and context information, which would cause the retrieval cue to only be compared with extralist images in the most extreme case because the images stored during study were not bound to context information.

Conclusion

For many purposes, common sense intuition is sufficient to guide scientific inference, but as knowledge in a field increases, explanations become more sophisticated and formalized. One cost associated with such an evolution is the decreasing utility of intuition. Even simple models often have nonintuitive consequences, so that their adequacy as an explanatory device needs to be assessed through analysis or simulation. The research presented here provides one small example: Data that on the surface seemed to imply a dual-process model (e.g., Balota et al, 2002; Hirshman et al, 2002; Joordens & Hockley, 2000; Reder et al, 2000) turn out to be compatible with a simpler single-process model (Shiffrin & Steyvers, 1997), a fact that needs to be taken into account when designing future research or drawing conclusions about memory processes and differences among populations. The demonstration here is reminiscent of a demonstration in another setting by Nosofsky and Zaki (1998). They showed that an extant exemplar model of categorization could, by varying one parameter value, account for the memory and categorization performance of memory-impaired individuals, data that seemed on the surface to call for a more complex dual-process model.

We have shown that there are a number of different ways that the single-process REM model can account for Hirshman et al.'s (2002) findings. The model that assumes noisier storage is by far the simplest of those we have explored (see Shiffrin & Steyvers, 1997, 1998, to compare the complexity of the various REM models of recognition). Of course, the fact that a single-process model can predict the results does not imply that a dual-process model is incorrect. In fact, a dual-process model contains a single-process model as a special case and therefore can predict every outcome achievable by its contained and restricted model. REM, in particular, contains separate familiarity and recall retrieval processes (e.g., Diller, Nobel, & Shiffrin, 2001; Shiffrin & Steyvers, 1998) and allows for both to operate during recognition tasks (Malmberg, Holden, & Shiffrin, 2004). In tasks requiring more complex judgments than simple recognition, evidence can be found for the use of dual processes (e.g., Malmberg et al., 2004). It is also important to note that it is possible that dual processes are used in simpler tasks, whether or not the evidence is sufficient to demonstrate this. Finally, we note that the hypothesis that memory impairments are due to a decrease in accuracy of storage rather than amount of storage is an interesting idea in its own right and is deserving of further research.

---

[2] The hypothesis that aging causes less accurate or more noisy cognitive processing has a long history in the area of developmental psychology. Gregory (1959), and later Layton (1975), proposed that an increase in noise was a significant factor contributing to the negative effect of aging on cognition. Although these models were mostly applied to account for deficits in selective attention, the same argument can be made that age-related deficits in memory are attributable to an increase in the amount of noise stored during study.

[3] Specifically, (a) the aging manipulation was, of course, between subjects (leading to the distinct possibility that the number of parameters of the model would be greater than the number of data points); (b) different populations demonstrated drastically different biases to respond "yes"; and (c) the aging manipulation was confounded with differences in list length.

References

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology, Vol. 1, Learning, memory, and thinking* (pp. 243–293). San Francisco: Freeman.

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). Memory for the infrequent in young, old, and early stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language, 46,* 199–226.

Buzsaki, G. (1989). Two-stage model of memory trace formation: A role for "noisy" brain states. *Neuroscience, 31,* 551–570.

Caplan, J. B., Madsen, J. R., Raghavachari, S., & Kahana, M. J. (2001). Distinct patterns of brain oscillations underlie two basic parameters of human maze learning. *Journal of Neurophysiology, 86,* 368–380.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of a pattern-analyzing skill in amnesia: Dissociation of knowing how and know that. *Science, 210,* 207–210.

Deadwyler, S. A., West, M., & Lynch, G. (1979). Activity of dentate granule cells during learning: Differentiation of perforant path input. *Brain Research, 169,* 29–43.

Dennis, S., & Humphries, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478.

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC–REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 414–435.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533.

Donaldson, W., MacKenzie, T. M., & Underhill, C. F. (1996). A comparison of recollective memory and source monitoring. *Psychonomic Bulletin & Review, 3,* 486–490.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Gardiner, J. M. (1988). Functional aspects or recollective experience. *Memory & Cognition, 16,* 309–313.

Gardiner, J. M., & Gregg, V. H. (1997). Recognition memory with little or no remembering: Implications for a detection model. *Psychonomic Bulletin & Review, 4,* 474–479.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18,* 23–30.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 12,* 8–20.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Gregory, R. L. (1959). Increase in "neurological noise" as a factor in aging. *Proceedings of the 4th International Congress on Gerontology, 1,* 314–324.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hirshman, E., Fisher, J., Henthorn, T., Arndt, J., & Passannante, A. (2002). Midazolam amnesia and dual-process models of the word-frequency mirror effect. *Journal of Memory and Language, 47,* 499–516.

Hirshman, E., & Henzler, A. (1998) The role of decision processes in conscious memory. *Psychological Science, 9,* 61–64.

Hirshman, E., & Master, S. (1997) Modeling the conscious correlates of recognition memory: Reflections on the remember–know paradigm. *Memory & Cognition, 25,* 345–352.

Hirshman, E., Passannante, A., & Henzler, A. (1999). The influence of Midazolam on implicit memory tests. *Brain & Cognition, 41,* 351–364.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory of episodic, semantic, and procedural tasks. *Psychological Review, 96,* 208–233.

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1534–1555.

Kintsch, W. (1967). Memory and decision aspects of recognition learning. *Psychological Review, 74,* 496–504.

Layton, B. (1975). Perceptual noise and aging. *Psychological Bulletin, 82,* 875–883.

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Effects of repetitions, similarity, and normative word frequency on old–new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 319–331.

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 630–644.

Malmberg, K. J., & Shiffrin, R. M. (in press). The "one-shot" hypothesis and the list-strength effect for free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition, 30,* 607–613.

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

Marr, D. (1971). Simple memory: A theory for the archicortex. *Proceedings of the Royal Society, London, 262,* 23–81.

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

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100,* 183–203.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2), *Psychological Review, 104,* 839–862.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247–255.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map.* Oxford, England: Clarendon University Press.

Polster, M., McCarthy, R., O'Sullivan, G., Gray, P., & Park, G. (1993). Midazolam-induced amnesia: Implications for the implicit/explicit memory distinction. *Brain & Cognition, 22,* 244–265.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 294–320.

Schulman, A. I. (1967). Word length and rarity in recognition memory. *Psychonomic Science, 2,* 47–52.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20,* 11–12.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163.

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

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from

memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73–95). London: Oxford University Press.

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

Tulving, E. (1983). *Elements of Episodic Memory*. New York: Oxford University Press.

Vizi, E. S., & Kiss. K. P. (1998). Neurochemistry and pharmacology of the major hippocampal transmitter systems: Synaptic and nonsynaptic interactions. *Hippocampus, 8*, 566–607.

Warrington, E. K., & Weiskrantz, L. (1970, November 14). Amnesic syndrome: Consolidation or retrieval? *Nature, 228*, 628–630.

Wilson, R. S., Bacon, L. D., Kramer, R. L., Fox, J. H., & Kaszniak, A. W. (1983). Word frequency effect and recognition memory in dementia of the Alzheimer type. *Journal of Clinical Neuropsychology, 5*, 97–104.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1341–1354.

Received March 3, 2003
Revision received July 25, 2003
Accepted July 29, 2003 ■