Psychological Review
2001, Vol. 108, No. 2, 452–478

0033-295X/01/$5.00 DOI: 10.1037//0033-295X.108.2.452

A Context Noise Model of Episodic Word Recognition

Simon Dennis and Michael S. Humphreys
University of Queensland

Item noise models of recognition assert that interference at retrieval is generated by the words from the study list. Context noise models of recognition assert that interference at retrieval is generated by the contexts in which the test word has appeared. The authors introduce the bind cue decide model of episodic memory, a Bayesian context noise model, and demonstrate how it can account for data from the item noise and dual-processing approaches to recognition memory. From the item noise perspective, list strength and list length effects, the mirror effect for word frequency and concreteness, and the effects of the similarity of other words in a list are considered. From the dual-processing perspective, process dissociation data on the effects of length, temporal separation of lists, strength, and diagnosticity of context are examined. The authors conclude that the context noise approach to recognition is a viable alternative to existing approaches.

Episodic recognition refers to the task of identifying a stimulus as having occurred within a particular episode or context. In a typical recognition experiment, participants process a study list of words and are then presented with a test list containing some old words from the study list and some new words, which did not appear. The participant's task is to determine which of the test words were presented at study. This basic design can be elaborated in a number of ways by including additional study lists and requiring participants to recognize words from individual lists or from all of the lists.

The recent literature on episodic recognition has been dominated by the *dual-processing* approach, especially as it is embodied within the process dissociation procedure, and the *item noise* approach, as it is embodied within global matching models. The dual-processing approach assumes that recognition involves *familiarity,* which is often assumed to be a context-insensitive automatic process, and *recollection,* a context-sensitive strategic process. The item noise approach assumes that recognition involves a single context-sensitive process in which noise is generated primarily from the other words in the study list.

Although they address the same recognition phenomena, the dual-processing and item noise approaches have differed markedly both in the data sets they have considered and the nature of the models they have proposed. The dual-processing approach has focused on manipulating study and test trial processing, typically in list discrimination designs, whereas the item noise approach has focused on differences in materials (such as word frequency and concreteness), associative information, and length and strength manipulations, typically within single-list designs. The dual-processing approach has involved the use of *measurement* models designed to estimate the contributions of automatic and controlled processing to the recognition decision, whereas the item noise approach has postulated *process* models describing specific representations (bindings), cues, and decision mechanisms occurring in recognition memory.

There has been an extensive literature critiquing both approaches (e.g., Clark, 1999; Clark & Gronlund, 1996; Dennis & Humphreys, 1998; Dodson & Johnson, 1996; Gruppuso, Lindsay, & Kelly, 1997; Humphreys, Pike, Bain, & Tehan, 1989; Mulligan & Hirshman, 1997; Ratcliff, Van Zandt, & McKoon, 1995). In this article, we draw on these critiques to integrate explanations from both approaches into the *bind cue decide model of episodic memory* (BCDMEM).

BCDMEM assumes that word recognition is a *context noise* process that involves cuing with a word to retrieve the set of contexts in which that word has been encountered. Performance is determined primarily by the other contexts in which the word has appeared and the degree of overlap between the study context and the context that the participant reinstates at test. On logical grounds, cuing with the context to retrieve the item and cuing with the item to retrieve the context are alternative bases for episodic recognition (J. A. Anderson & Bower, 1972; Humphreys, Wiles, & Dennis, 1994).[1] By introducing a context noise model, we provide

Simon Dennis and Michael S. Humphreys, Key Centre for Human Factors and Applied Cognitive Psychology, University of Queensland, Brisbane, Queensland, Australia.

This research was supported by Australian Research Council Grants A79701517 and A79532361. We would like to thank Scott Boland and Simon Finnigan for their assistance with the experiments. We would also like to thank Kerry Chalmers for allowing us to use her data set and Glen Smith for comments on earlier versions.

Correspondence concerning this article should be addressed to Simon Dennis, Key Centre for Human Factors and Applied Cognitive Psychology, University of Queensland, Brisbane, Queensland 4072, Australia. Electronic mail may be sent to s.dennis@humanfactors.uq.edu.au.

[1] The recognition portions of J. A. Anderson and Bower (1972) and BCDMEM are generally similar. In addition to cuing with the word to retrieve prior contexts, both treat words as single nodes and context as a collection of abstract features. Furthermore, BCDMEM's Bayesian decision process is relatively similar to the feature matching process used by Anderson and Bower. However, further developments in the Anderson and Bower framework took a different tack. Anderson and Bower (1974) used interpretable features (e.g., list tags). Current models arising from this tradition also incorporate item noise (activation spreads from a list node) as well as context noise (J. R. Anderson, Bothell, Lebiere, & Mantessa, 1998). In addition, the BCDMEM recall model, which is only briefly mentioned in this article, differs substantially (e.g., recalled words are represented as a collection of features).

a test bed for comparing the strengths and weaknesses of these approaches. Furthermore, BCDMEM demonstrates how certain important intuitions from the dual-processing perspective can be incorporated into a process model.

In this article, we focus on recognition, but a significant portion of the memory modeling literature has focused on the differences between recognition and recall. We believe that many of these differences can be parsimoniously explained by assuming that recognition is a context noise process and recall is an item noise process. At various stages throughout the text, we demonstrate the utility of this assumption.

We begin by describing BCDMEM. We then consider the item noise and dual-processing approaches in turn, outlining their primary assumptions and demonstrating how BCDMEM is able to account for data from each approach. BCDMEM makes unique predictions in multilist paradigms involving the manipulation of processing and temporal similarity. We discuss data from these paradigms and conclude by discussing the different domains of application of the three approaches and directions for further research.

The BCDMEM

As the name suggests, there are three critical components of the BCDMEM: the binding mechanism, the cues used, and the decision rule. The binding mechanism specifies how elements of an episode including words, contexts, and other information are bound in episodic memory. The cues are the elements that are used to initiate retrieval. Not all of the information available to an individuals needs to be used as a cue, so part of the theory involves specifying which components are used as cues in a given experimental paradigm. Finally, the decision rule takes the results of retrieval and outputs the required information in the form of a word in recall paradigms or a yes–no decision in recognition paradigms. A complete theory of episodic memory must address all of these components. However, in explicating the recognition model, only the cues and decision mechanism are critical, so in the following section we focus on these components.

The Mechanism

Figure 1 outlines the components of the architecture of BCDMEM relevant to the recognition study trial. Active nodes (activation value equals 1) are represented by solid circles, and inactive nodes (activation value equals 0) are represented by open circles. At the input layer, words are represented as individual nodes (local codes). At the output layer, the study trial context is represented as a pattern of activity (distributed code with sparsity s and length v). Each node at the input layer is connected through associative weights to the output layer. At study, the current word is instantiated on the input layer, and the current context pattern is instantiated on the output layer. Learning occurs by setting the weight connecting an input node to an output node to one with probability r (learning rate) whenever the input and output nodes are both active.

Figure 2 outlines the components relevant to the recognition test. Here the word node is reinstated at the input layer, which in turn retrieves a composite vector at the output layer. This composite contains the contexts with which the word has been associated. For a target word, the output layer will contain some of the study context (the amount depends on the learning rate). Regardless of whether a word is a target or a distractor, it will have appeared in nonstudy contexts, so there will be weights that were learned during those episodes. These weights will in turn activate nodes at the output layer. The probability that an output unit is active in the retrieved context vector, as a result of previous learning, is called context noise (p). Because the retrieved context vector is a composite of many different study contexts, p is much larger than s. Conceptually, the probability p represents the balance between past learning and forgetting. That is, we would expect p to monotonically increase with the frequency or number of contexts in which a word has been encountered and to monotonically decrease as the time from the last encounter increases. However, we would not necessarily expect rapid changes in p as a function of either frequency or recency (e.g., it might take a period of several years without encountering a word before p declines by a significant amount).

The retrieved vector is then compared against the reinstated study context vector to determine whether the context vector for which the participant is looking has been associated with the current test word. There may be components of the study context that are retrieved even when a distractor is presented as a consequence of the overlap between the study context and the contexts in which the distractor has been encountered. Likewise, when a target is presented, the output layer may be missing components of the study context because they were not learned. As a consequence, the retrieved context vector will not be identical to the reinstated vector, and errors may arise.

BCDMEM assumes that the mechanism underpinning recognition decisions approximates the optimal decision rule in a Bayesian sense (cf. J. R. Anderson & Milson, 1989; Glanzer & Adams,

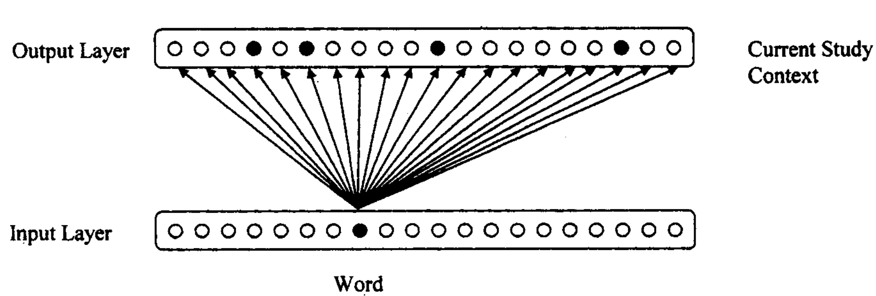


Figure 1. Bind cue decide model of episodic memory at study.

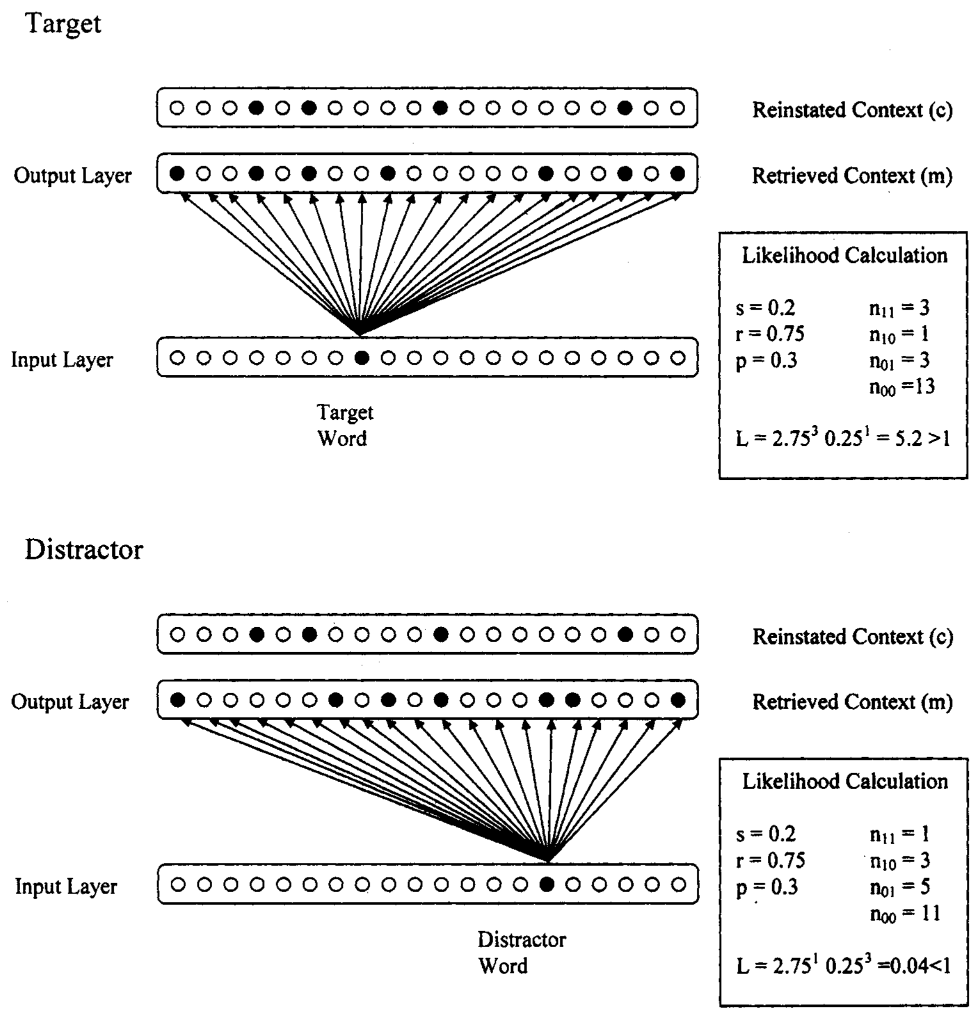


Figure 2. Bind cue decide model of episodic memory at test.

1990; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). Consequently, the decision can be characterized as the odds ratio (Equation 1):

$$\frac{p(\text{old}|\text{data})}{p(\text{new}|\text{data})} = \frac{p(\text{old})}{p(\text{new})}\frac{p(\text{data}|\text{old})}{p(\text{data}|\text{new})}. \qquad (1)$$

The data referred to in the odds expression are the reinstated context vector (**c**) and the retrieved context vector (**m**). The odds ratio depends on how well these vectors match. The word will be considered old if the probability that it is old given the data is greater than the probability that it is new given the data (i.e., if the odds ratio exceeds one).

In many experiments participants see equal numbers of targets and distractors, so it could be assumed that, in the absence of specific manipulations of the criterion, $P(\text{old}) = P(\text{new}) = .5$, and the prior probabilities cancel. In this case, the odds ratio is equal to the likelihood ratio, that is, $P(\text{data}|\text{old})/P(\text{data}|\text{new})$.

Because both the reinstated context vector and the retrieved context vector are binary, there are four types of match (i.e., $c_i = 1$ and $m_i = 1$, $c_i = 1$ and $m_i = 0$, $c_i = 0$ and $m_i = 1$, and $c_i = 0$ and $m_i = 0$). The probability of a given sort of match is independent of the component in which that match occurs, so the data can be summarized by the numbers of matches of each kind. Using the convention that the first component is the reinstated context and the second component is the retrieved context, let n_{11} be the number of 11 matches, n_{10} the number of 10 matches, n_{01} the number of 01 matches, and n_{00} the number of 00 matches (note that vector length $v = n_{00} + n_{01} + n_{10} + n_{11}$):

$$\begin{aligned} P(\text{data}|\text{old}) &= \Pi_i P(c_i \ \& \ m_i|\text{old}) \\ &= P(c_i = 1 \ \& \ m_i = 1|\text{old})^{n11} P(c_i = 0 \ \& \ m_i = 0|\text{old})^{n00} \\ &\quad P(c_i = 0 \ \& \ m_i = 1|\text{old})^{n01} P(c_i = 1 \ \& \ m_i = 0|\text{old})^{n10} \\ &= [P(c_i = 1|\text{old})P(m_i = 1|c_i = 1 \ \& \ \text{old})]^{n11} \\ &\quad [P(c_i = 0|\text{old})P(m_i = 0|c_i = 0 \ \& \ \text{old})]^{n00} \\ &\quad [P(c_i = 0|\text{old})P(m_i = 1|c_i = 0 \ \& \ \text{old})]^{n01} \\ &\quad [P(c_i = 1|\text{old})P(m_i = 0|c_i = 1 \ \& \ \text{old})]^{n10} \qquad (2) \end{aligned}$$

A similar equation can be written for $P(\text{data}|\text{new})$.

We can now restate the likelihood ratio in terms of the parameters of the model that have been introduced to this point. Sparsity (s) is the probability that a component of a study context vector is equal to one. Learning (r) is the probability that the link between a word node and an active context component is learned during study. Context noise (p) is the probability that a component of the retrieved vector is equal to one because of extra-experimental contexts in which the word has been seen. Vector dimensionality (v) is the length of the reinstated and retrieved vectors. Substituting into the previous equations, we get Equations 3–5.

$$P(\text{data}|\text{old}) = [s(r + p - rp)]^{n11}[(1 - s)(1 - p)]^{n00} [s(1 - r)(1 - p)]^{n10}[(1 - s)p]^{n01}. \quad (3)$$

$$P(\text{data}|\text{new}) = [sp]^{n11}[(1 - s)(1 - p)]^{n00} [s(1 - p)]^{n10}[(1 - s)p]^{n01}. \quad (4)$$

Thus,

$$P(\text{data}|\text{old})/P(\text{data}|\text{new}) = [(r + p - rp)/p]^{n11}(1 - r)^{n10}. \quad (5)$$

Note that in this simple version of the model the number of 01 and 00 matches has no impact on the likelihood ratio, because the terms in the numerator and denominator that are raised to these powers are the same. Because learning cannot occur when the component in the study context vector is zero, the 01 and 00 matches cannot help to distinguish new and old words.

As mentioned previously, when there is no specific manipulation of the criterion, it is assumed that a word will be called old if the probability that it is old given the data is greater than the probability that it is new given the data, which is true when the likelihood ratio is greater than one. In general, then, as the mean likelihood ratio approaches one, from above in the case of targets and from below in the case of distractors, we expect performance to degrade. We can begin to understand how the preceding likelihood function simulates performance by looking at how its expected value varies as a function of the parameters (note that a full exposition would consider the complete likelihood distribution). First, as context noise (p), which represents word frequency, approaches one, $(r + p - rp)/p$ approaches one and the expected value of n_{10} approaches zero, so the expected value of $P(\text{data}|\text{old})/P(\text{data}|\text{new})$ approaches one. In other words, performance decreases as word frequency increases. Second, as learning (r) approaches zero, $(r + p - rp)/p$ approaches one and $1 - r$ approaches one, so the expected value of $P(\text{data}|\text{old})/P(\text{data}|\text{new})$ approaches one. Thus, as study time or number of repetitions decreases, so does performance.

A numerical example. To solidify understanding of the model, we now work through three simple numerical examples. First, consider the network in Figure 2. In constructing this example, vector length was set to 20, sparsity was set to 0.2, context noise was set to 0.3, and learning rate was set to 0.75. Four of the reinstated context units were active, six of the retrieved context units were active, and, for the target, three of the retrieved units corresponding to the four units in the reinstated context were active. Figure 2 shows the number of matches of each type and the likelihood calculation. With these parameters, discrimination is very good. The likelihood ratio for the target is 5.2, which is well above 1, and hence this word would have been identified as old. The likelihood ratio for the distractor is 0.04, which is well below 1, and so this word would have been correctly classified as new.

Now consider the effect of increasing the context noise to 0.6 (modeling an increase in the frequency of the word; see Figure 3). Twelve of the retrieved context vector units were active, and the target likelihood was 5.06, whereas the distractor likelihood was 0.84. Thus, the target and distractor were still correctly classified, but the matching values were closer to one. Noise in the match is more likely to generate a false alarm or a miss. Note also that the n_{11} and n_{10} matches are equal in the original target and the high-frequency distractor, yet in the first case the model responds yes, whereas in the second case the model responds no. These response changes occur as a consequence of the change in the ratio of the probabilities that n_{11} matches will occur. For example, the ratio of the probability that an n_{11} match will occur given a target to the probability that it will occur given a distractor is $(r + p - rp)/p$. These ratios of probabilities are referred to as the weighting of the matches. In the original case the weighting of the n_{11} matches was 2.75, whereas in the high-frequency case the weighting was only 1.5. The model relied on this change in the weighting to correctly classify the words. However, note the effect that the change in context noise had on the n_{01} and n_{00} matches. These changes allowed the estimation of the different weighting parameters.

Finally, consider the effect of decreasing the amount of learning to 0.5 (see Figure 4). In terms of the number of matches, the distractor did not change. However, for the target, one of the 11 matches became a 10 match. The resultant likelihood ratios were 1.18 for the target and 0.27 for the distractor. Again, discriminability was reduced.

Adding the contextual reinstatement parameter. In the derivations just outlined, it was assumed that the ability to retrieve or otherwise reconstruct the study context at test (contextual reinstatement) is perfect. The context used at test is identical to that used at study. It seems more likely, however, that as a consequence of factors such as delay, features of the original context vectors will be lost. (Note that, in future instantiations of the model, it may be necessary to consider the possibility that spurious features will arise during the reinstatement process. We have chosen to include only the loss of units at this stage for mathematical simplicity.) The contextual reinstatement parameter (d) is the probability that a unit that was a one in the study context will be a zero in the reinstated context. The likelihood ratio can be rederived taking into account contextual reinstatement:

$$\begin{aligned} &P(c_i' \,\&\, m_i|\text{old})/P(c_i' \,\&\, m_i|\text{new}) \\ &= \{[1 - s + d(1 - r)s]/[1 - s + ds]\}^{n00}(1 - r)^{n10} \\ &\{[p(1 - s) + d(r + p - rp)s]/[p(1 - s) + dps]\}^{n01} \\ &[(r + p - rp)/p]^{n11}, \end{aligned} \quad (6)$$

where c' is the reinstated context vector (i.e., the original context vector minus the ones that have been lost in the process of reinstatement; see Appendix A for the derivation).

In this version of the model, the 00 and 01 matches do not cancel because a zero in the reinstated context does not preclude learning

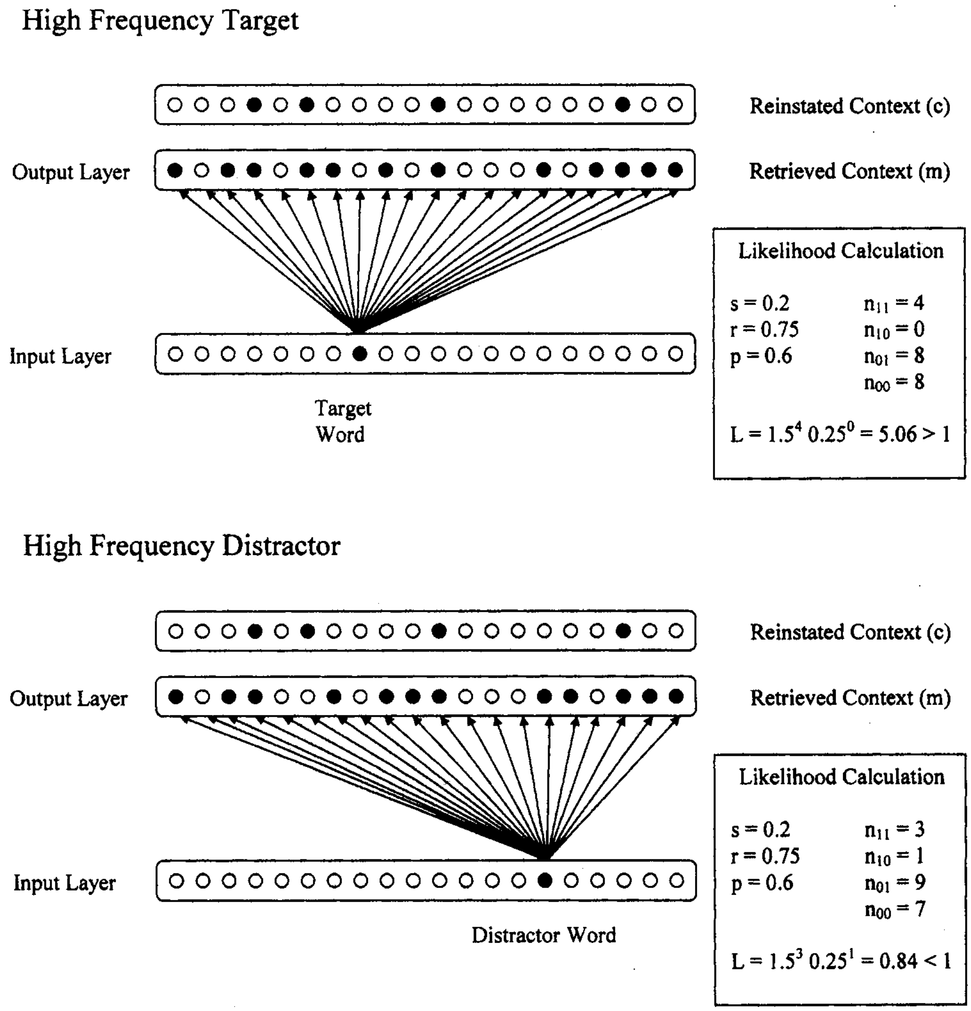


Figure 3. Numerical example showing the effect of increasing context noise.

at study. However, if d is set to zero, indicating that reinstatement is perfect, the likelihood ratio reduces to that previously derived, and the 00 and 01 matches cancel.

About the parameters. In Bayesian models of this kind, the parameters play two different roles. First, they describe the encoding and retrieval processes of the model. For instance, the learning rate parameter affects the probability that an association between an input node and an output node will be learned at study. Second, they affect the decision mechanism. In the first case, there is no problem. The parameters of the model are related to parameters of the physical system. In the second case, however, the decision mechanism, which may well be physically distinct from the storage system, must have access to estimates of these parameters. That is, it must be capable of calculating the ratios of the probabilities that each type of match will occur. In this case, it becomes important to ask how the decision mechanism is able to derive or learn the parameters to calculate the ratios.

First, it should be noted that the sparsity and vector length parameters are not affected by experimental manipulations. The sparsity parameter was set to 2%, and the same value was used in all of the simulations. Preliminary investigations showed that a vector dimensionality between 200 and 1,000 provides hit rates and false alarm rates in the appropriate ranges for recognition memory experiments. A length of 200 was used in the initial simulations and was extended to 1,000 when the overlap between vectors had to be manipulated.

Second, except in the case of context noise, the setting of the decision parameters requires processes of the same kind as those involved in the setting of a criterion in the global matching models. It is commonly assumed that participants are capable of altering criteria as a consequence of instructions or when the proportion of old and new words changes (Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995). Parameters such as learning (r) and contextual reinstatement (d) that refer to the entire list could be set on a similar basis. Note that when within-list manipulations of learning rate are simulated, an average value is used in the decision rule. It is not necessary for the strength to be estimated on a word-by-word basis.

Over a large number of experimental manipulations, increases in hit rates are accompanied by decreases in false alarm rates. This pattern of results is called the mirror effect, and it has been argued that such a pattern is ubiquitous in recognition memory (Glanzer &

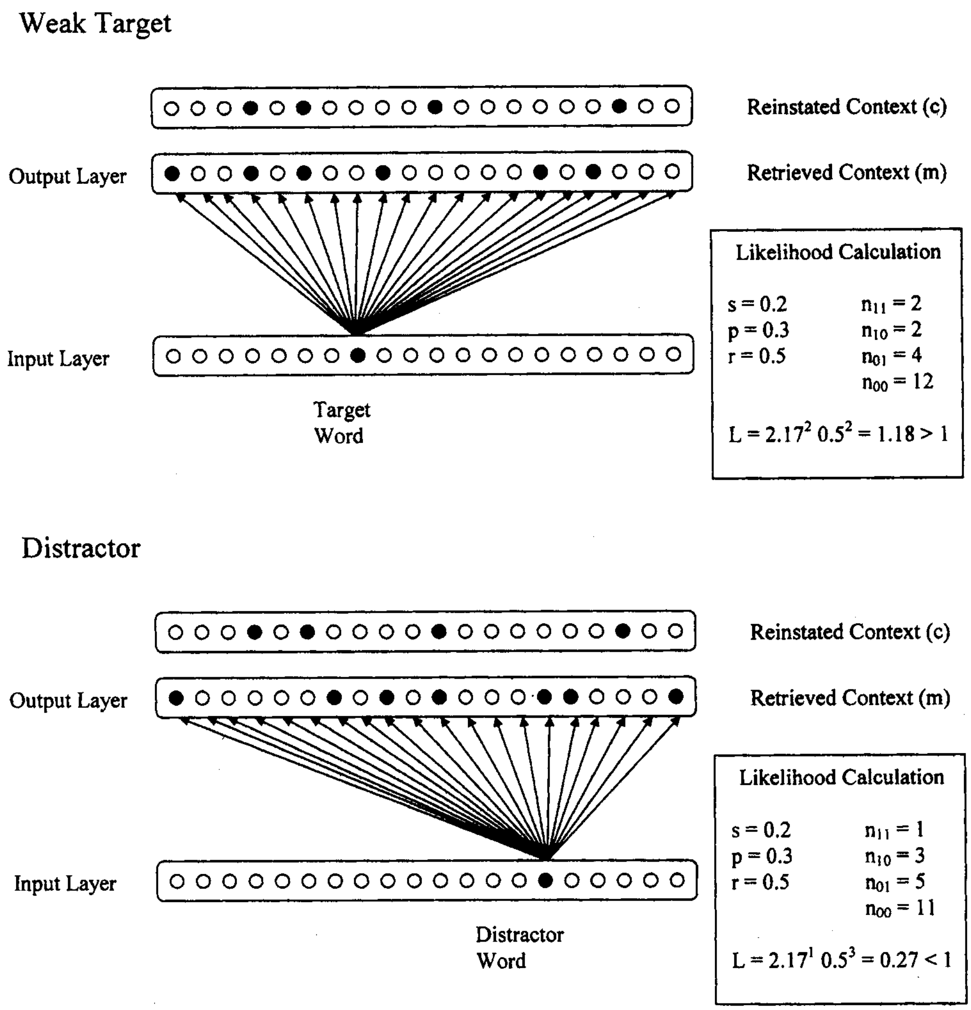


Figure 4. Numerical example showing the effect of decreasing learning.

Adams, 1990; Glanzer & Bowles, 1976; McClelland & Chappell, 1998). This ubiquity provides much of the motivation behind Bayesian models such as attention–likelihood theory (Glanzer & Adams, 1990), BCDMEM, retrieving effectively from memory (REM; Shiffrin & Steyvers, 1997), and the model of McClelland and Chappell (M&C model; 1998), because the mirror effect follows directly from the decision mechanisms. By contrast, in global matching models, a yes–no criterion is introduced as a completely free parameter, and mirroring occurs as a consequence of the fitting process but is not a necessary consequence of the decision mechanism.

Unlike the learning and contextual reinstatement parameters, the context noise parameter cannot be set on a list-wide basis, because effects of word frequency persist even when mixed lists are used (Glanzer & Adams, 1990). Furthermore, as the earlier numerical example demonstrated, mirroring as a function of context noise depends on the use of the appropriate context noise parameter in the decision mechanism. However, the context noise parameter can be approximated directly from the information available in the retrieved and reinstated context vectors. Consider the n_{00} matches (Equations 7–9):

$$P(c_i = 0 \;\&\; m_i = 0|\text{old}) = P(c_i = 0 \;\&\; m_i = 0|\text{new}) \quad (7)$$
$$= (1 - s)(1 - p).$$

Now,

$$n_{00} = vP(c_i = 0 \;\&\; m_i = 0|\text{old}) = v(1 - s)(1 - p). \quad (8)$$

Thus,

$$p = 1 - n_{00}/[v(1 - s)]. \quad (9)$$

Because v and s are fixed parameters, the system can, in principle, estimate context noise from the information available in a match on a word-by-word basis. That the requisite information is available directly from the retrieved vectors makes it plausible that a low-level mechanism could be constructed to extract the context noise parameter automatically.

Context and Its Reinstatement

In a model such as BCDMEM, in which context plays a pivotal role, it is important to outline what information the context vectors might represent. Our contention is that at least two distinct forms of context, processing and temporal, are critical to episodic recognition.

Our concept of processing context is generally similar to the dual-processing concept that is used in the recollection process. That is, processing context is descriptive of the actions used during processing. Was the task to count the number of vowels in the word, to solve an anagram, or to make a rating in regard to pleasantness? Several studies have demonstrated that the diagnosticity of the study task is important in list discrimination (Dodson & Johnson, 1996; Gruppuso et al., 1997; Mulligan & Hirshman, 1997), suggesting that type of processing is a critical component of the context representation. However, our features are subsymbolic (Smolensky, 1988; they do not have any interpretation or meaning in their own right), whereas the output of the recollection process appears to be meaningful or symbolic.

In addition, the features in the reinstated context must be representative of the types of features that could be retrieved for an old word, not features that are specific to a particular old word. One way this could occur is if the system "knows" that features in one part of the context vector are likely to be active if the word was read and other features are likely to be active if the word was heard. Another, not necessarily exclusive possibility is that context is closely linked to participants' representation of the task they are to perform during the study session. If participants are to remain on task during study, they must have an invariant representation of the task. In addition, they may have representations of components of the task, such as the operation of the buttons in a pleasantness-rating task. Processing context may include a summary of the instructions about the task provided by the experimenter or a task description established by the participant in response to those instructions.

Temporal context is a time-varying type of context. In BCDMEM, we assume that a form of contextual drift occurs (cf. Howard & Kahana, 1998, 1999). As time progresses, new features become active, and old features become inactive. Consequently, the overlap between temporal context vectors representing two lists will be greater if the lists occur closer in time. BCDMEM also assumes that temporal context can change from the start to the end of a list, providing a basis for discrimination (in contrast with REM, in which context is assumed to be constant for the duration of the list; Shiffrin & Steyvers, 1997). Allowing context to change on this time scale permits an explanation of paradigms such as that of Hall (1996, Experiment 2). Hall (1996) used a list discrimination design, but instead of being required to discriminate two separate lists, participants were required to discriminate list halves. That is, they were required to say "yes" to words from the second half of the list and "no" to words from the first half. Because participants were unaware of the length of the list at study, it was not possible for them to identify which half they were working with at that time. Nevertheless, participants were capable of making list half discriminations. We assume that Hall's participants were using an end of list context. A strong match to the end of list context would indicate that the word appeared in the second half. A weak match would indicate that it appeared in the first half.

Howard and Kahana (1999) have proposed an explanation for rapid within-list context changes that is very compatible with BCDMEM. They assume that as each study word is presented, it is used to retrieve the previous contexts associated with that word. The retrieved contexts are then incorporated into the evolving list context. With this assumption, they were able to model the serial position of the first word recalled in free recall. With the additional assumption that each test word is used to retrieve its associated context, they could also model the lag function relating the serial position of the kth word recalled to the serial position of the $k -$ 1st word recalled.

Having distinguished between processing and temporal forms of context, we are now in a somewhat better position to consider the question of how participants have some understanding about the nature of the to-be-retrieved information. The first observation to make is that when the test is administered shortly after the study list, it may not be necessary to reinstate context at all. The representation of processing and temporal contexts that has been associated with each study word may remain active and may be used directly in the comparison operation. Under these circumstances, we would expect the processing context to have remained stable during the presentation of the list, so the processing context used for comparison will be very similar to that present when the word was studied. Temporal context, however, may have drifted from the start to the end of the list (especially with long lists or multiple lists). Consequently, we anticipate that effects of length and lag will be pronounced when the test immediately follows study.

However, when there is a substantial delay filled with distracting activity between study and test, reinstatement of context will be necessary. Typically, test instructions refer to the nature of processing conducted at study. For instance, participants might be asked to say yes to the words that they rated for pleasantness. In this case, they could reconstruct the context from instructions in essentially the same fashion in which they originally established the processing context. Reinstating temporal context or a processing context not included in the test instructions would require memory retrieval. Just as we assume that the presentation of list words involves the formation of a word-to-context association, we also assume that presentation of study instructions involves the formation of an instructions-to-context association. Using aspects of the instructions as a cue, participants may retrieve the context vector in the same fashion that they retrieve any other memory. Once retrieved, the context vector would be available for comparison against the context vectors retrieved for list and nonlist words.

We also anticipate that, under some conditions, the reinstated context vector encompasses a combination of study context vectors. For instance, in the inclusion condition of a dual-list design, we might expect that the participant would form a reinstated context that incorporates the context vectors from both lists. In BCDMEM, we model the experiment-wide context by taking the bitwise odds ratio of the contexts for each study list. That is, the experiment-wide context will contain a one whenever either of the two study list contexts contains a one in that position in the vector. The resulting reinstated context will contain more ones (i.e., be more diffuse) than either of the study list contexts, and diffuseness will increase if the lists are separated in time (so that the amount of overlap decreases).

Local Codes and Item Novelty

The majority of recognition experiments involve well-known words as stimuli. In BCDMEM, it is assumed that the high amount of learning that is likely to have occurred with these items will have the effect of orthogonalizing their representations within the episodic binding system (cf. McClelland, McNaughton, & O'Reilly, 1995). We model this through our assumption that each word activates a different input node (a local code) even if those words are synonyms, antonyms, associates, or category members or if they are physically similar.

The exception to this rule arises with morphemic relatives, which we assume have sufficiently similar representations that they may be captured by the same node in the binding mechanism. This assumption provides a relatively straightforward explanation for the thresholdlike effects with morphemic relatives reported by Hintzman, Curran, and Oppy (1992). That is, aside from some minor effects due to the probabilistic learning mechanism, we would not expect repeated presentations of a target word to increase the probability of falsely recognizing a morphemic relative. This lack of increase in the probability of recognition occurs because the probability of being captured by a node is controlled by the similarity of the representations, which is not changed during study. However, we would expect that frequency judgments would increase monotonically with the number of presentations, given that the lure was falsely recognized.

With words that the participant has not previously encountered, we expect that the local code assumption will be inadequate. First, some novel words (e.g., nonsense words) have an obvious component structure. With these stimuli, a participant could encode the word by selecting a subset of its letters. With partial encodings, it is possible that a distractor that is similar to a target (it overlaps in letters) will be encoded identically. Such a process could explain the high false alarm rate Postman (1951) found with nonsense words.

Second, in some situations words may be recognized not because a node in the binding mechanism has been associated with a context but because a lexical representation for that word exists. For example, Maddox and Estes (1997) used a three-phase design to investigate the recognition of digit and letter trigrams and pronounceable nonwords. In the first, or familiarization, phase, the words were presented a varying number of times. The second and third phases were the conventional study and test phases of a recognition experiment. Maddox and Estes found that previous familiarization increased both the hit rate and the false alarm rate. These changes in hits and false alarms resulted in the ability to discriminate between words that had and had not been present remaining the same or decreasing. Chalmers and Humphreys (1998) also used a three-phase design with very-low-frequency words, which would have been nonsense words to the vast majority of their participants. During the familiarization phase, these words were trained either with or without their definitions. When they were trained without their definitions, Maddox and Estes's (1997) finding of a decrease in ability to discriminate was replicated. However, for words trained with their definitions, there was an increased ability to discriminate.

Chalmers and Humphreys (1998) then looked at the effect of familiarization on recency and frequency judgments. One week after the familiarization phase, participants received two study lists separated by 24 hr with a test shortly after the second study list. From the perspective of the participants, words were studied once or three times on the previous day or once or three times on the present day. The differential effect on frequency and recency judgments of training with and without the definitions can be seen most readily in judgments for words studied on the present day. With frequency judgments, previous familiarization increased the probability that participants would judge that the present day's words had occurred three times. This held for both words studied with their definitions and words studied without their definitions. In contrast, with recency judgments, the present day's words were judged more accurately when the previous familiarization had involved the word's definition and less accurately when it had not involved the word's definition. In addition, with words trained without their definitions, previous familiarization decreased the probability that a word would be judged as having occurred on the present day. Chalmers and Humphreys (1998) concluded that participants were more capable of forming an episodic memory if words had been trained with their definitions. However, in the absence of episodic memory, they assumed that participants used information about the existence of a representation differentially in making frequency and recency judgments.

In summary, when recognition involves novel stimuli or stimuli that are encoded in a componential fashion, an assumption of local codes is unlikely to be justified. In BCDMEM, however, we are attempting to capture recognition performance with well-learned word stimuli.

We have now described the components of BCDMEM. What remains is to determine how it accounts for important data from both the item noise approach and the dual-processing approach. We start with the item noise approach.

BCDMEM and the Item Noise Approach

Item noise models assume that an image, trace, or association is laid down at each study opportunity and that the recognition decision is based on the matches of the current word against the images of the words in the list (see Figure 5). Typically, new words match weakly with all of the study words, whereas old words match the other words weakly and their own image strongly. Consequently, the overall match for old words will tend to be higher than that for new words, and a criterion can be set above which the participant says yes and below which the participant says no. Because the other words from the list enter into the match, they generate most of the noise, and consequently item noise models are sensitive to manipulations of the other words in the list.

The prime examples of the item noise approach are the global matching models, including search of associative memory (SAM; Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981), the

Test Item ⊗ Study Item 1 =	Match 1
	+
Test Item ⊗ Study Item 2 =	Match 2
	+
...	...
	+
Test Item ⊗ Study Item N =	Match N
	+
Test Item ⊗ Pre-experimental Memories =	Pre-experimental Match
	=
	Global Match

Figure 5. The general item noise framework. The ⊗ symbol denotes the matching operation (e.g., dot product in the matrix model).

theory of distributed associative memory (TODAM; Murdock, 1982), the composite holographic associative recall model (Eich, 1982), the matrix model (Humphreys, Bain, & Pike, 1989; Pike, 1984), and Minerva II (Hintzman, 1984). Although these models differ in many respects (see Clark & Gronlund, 1996, for a review), their basic mechanisms involve summing the matches of cue and list images. More recently, a series of models have emerged that are able to account for data that were problematic for the global matching models. This class of models is based on a Bayesian conception of the recognition process and includes REM (Shiffrin & Steyvers, 1997) and the M&C model (McClelland & Chappell, 1998). Whereas M&C is a pure item noise model, REM includes both item noise and context noise features. However, in applications to date, the item noise aspects of REM have been emphasized.

For our purposes, there are three important attributes of the item noise models. First, they assume that, for the most part, recognition involves a single context-sensitive process. Second, this process is assumed to be subsymbolic. Third, REM and M&C assume that memory processes are optimized to the environment (cf. J. R. Anderson & Milson, 1989) and involve a Bayesian decision process. BCDMEM also involves a Bayesian subsymbolic, context-sensitive mechanism.

In this section, we demonstrate how BCDMEM can account for data that are important within the item noise approach. These data include the mirror effect for word frequency and concreteness and the null list strength effect. We also consider two challenges to BCDMEM and other context noise models posed by list length effects and the effects of word similarity.

List Strength and List Length

Critical to understanding whether recognition involves interference from the other words in a list or the other contexts in which the word has appeared is the effect that manipulating some words on a list has on the other words. Item noise accounts, such as the global matching models, predict that strengthening other words or adding new words will hurt performance because the variance of the matching strength will increase (Clark & Gronlund, 1996). In contrast, a model in which interference is generated by the other contexts in which the word has been seen does not predict either a list strength or a list length effect. As long as there is no overlap in the binding layer of the representations of the two words (this occurs in BCDMEM through the use of local representations), strengthening the association of another word to the study context does not affect the memory retrieved to a completely different word. Similarly, adding another word to the list does not affect the match of a completely different word.

A number of studies have demonstrated the lack of a list strength effect (Murnane & Shiffrin, 1991; Ratcliff, Clark, & Shiffrin, 1990). Whereas distributed composite models such as TODAM (in its original formulation) and the matrix model have had great difficulty accommodating this finding (see Murdock & Kahana, 1993a, 1993b; Orht & Gronlund, 1999; Shiffrin, Ratcliff, Murnane, & Nobel, 1993), it is possible to adjust local models such as SAM to predict no list strength effect by assuming that increasing strength also decreases similarity to other words (differentiation; Shiffrin, Ratcliff, & Clark, 1990). However, the adjustment requires a delicate balancing of similarity and strength parameters. More recent models such as REM (Shiffrin & Steyvers, 1997) and M&C (McClelland & Chappell, 1998) suggest that episodic word representations are initially similar and that strengthening involves filling in the features of the representation. The more features a representation has, the greater the opportunity for it to differ from other words. Such a mechanism provides a more principled explanation for why strength and similarity might be related in the appropriate fashion.

Whereas the revised versions of SAM, REM, and M&C are capable of producing a null list strength effect, they all propose that item noise is responsible for a length effect in recognition. The effect of length in recognition memory would appear to be one of the most established results in the memory literature. However, there are a number of variables closely associated with length that are also candidate explanations. No study has controlled for all of these variables, and most studies have left more than one uncontrolled.

The first candidate is retention interval. Clearly, if study list length is manipulated and the test is presented immediately after the list, then the average retention interval will be longer for the long list. Attempts to control for retention interval can be divided into those involving a retroactive design and those involving a proactive design. In the retroactive design, study–test lag is equated by filling the period after presentation of the study list with filler activity and comparing only the initial words in the long list with those in the short list. Using this design, Schulman (1974) found no effect of length in a forced choice test. Bowles and Glanzer (1983) also used the retroactive design. They did not analyze the retroactive length effect separately, but the mean size of the effect (proportion correct) was small (.033). In addition, in the third experiment of Murnane and Shiffrin (1991), in which a yes–no recognition test was used, the effect of length was not significant. In contrast to previous work, Gronlund and Elam (1994) did find significant effects using a retroactive design. Differences in d' values were 0.31 and 0.65 in their first and second experiments, respectively (a possible explanation is offered below).

In the proactive design, the time to the test from the end of the short and long lists is equated and testing occurs only on the final words in the long list. The effects of length in these conditions have typically been somewhat larger. In Bowles and Glanzer's (1983) study, the mean size of the effect (proportion correct) was .068. The overall effect of length when retroactive and proactive results were combined was significant, most probably driven by the proactive results. Similarly, Underwood (1978) used a forced choice test and found a significant effect of length of approximately the same magnitude. Orht and Gronlund (1999) found a much larger effect (0.95 on a d' scale). Underwood, citing the stability of word difficulty across list lengths and the lack of cumulative proactive interference results in other recognition paradigms, argued against direct proactive interference as the locus of the effect.

As a second candidate, Underwood (1978) suggested that the length effect in the proactive design might be a consequence of participants losing attention as they process the final words in the long list. In Bowles and Glanzer's (1983) study, the long list contained 240 words. In the Underwood (1978) study, the longest list contained 80 words, and in Ohrt and Gronlund's (1999) study, the long list contained 82 words. In all three cases, words were presented for a 1.5-s to 2.0-s interval under intentional learning instructions, but with no specific processing requirements and no way of ensuring that attention was maintained. Under these conditions, it seems plausible that attentional lapses could play a role, especially in the Ohrt and Gronlund (1999) study, in which participants were involved in four 50-min sessions.

A third candidate explanation is displaced rehearsals. In the retroactive design, participants may choose to spend part of the period between study and test rehearsing the items. To the extent that this occurs, it will favor the short list, because there is more time to rehearse and because in the long list the rehearsal is likely to be spread between the early (tested) and late (untested) items. Note that both experiments of Gronlund and Elam (1994) involved intentional conditions, which increased the likelihood of rehearsal of the short list.

A fourth candidate for the locus of the length effect is the contextual reinstatement process. In our earlier discussion of the reinstatement of context, we suggested that when the test occurs soon after study, participants may not reinstate a context at all. They may use the existing temporal context. In Gronlund and Elam's (1994) experiment, only 9 s of filler activity followed the long lists. Under these conditions, it is possible that the participants used an end of list context (note that this is the assumption made by Howard & Kahana, 1998, 1999, in modeling the serial position of the first word recalled in a free recall paradigm). Performance on the long list was measured on the first words in the list (because of the retroactive design). Consequently, the end of list context may not have been a good match for the context of words early in the list. In contrast, 69–70 s of filler activity followed the short lists, making it more likely that participants would be required to reinstate context. The reinstated context might have represented either a start of list context or a processing context. Both of these contexts might be expected to match better than the end of list context in the case of the long list, so performance would be better in the short list than the long. Furthermore, Dennis and Humphreys (1998) showed that by assuming that long lists compromised the reinstatement process and allowing the contextual reinstatement parameter (d) to vary between the short and long lists, the data set from Murnane and Shiffrin (1991, Experiment 3) can be modeled by BCDMEM.

A more diagnostic approach, however, is to ask whether there is still a list length effect when retention interval, attention, rehearsal, and contextual reinstatement controls are used. Item noise accounts propose that the length effect for both recognition and recall is a direct consequence of the other words in the list. If the list length effect is very small in recognition, then these models must predict that it will also be very small in recall. In contrast, BCDMEM can handle a very small effect in recognition and a much larger effect in recall because it cues with the word in recognition and the context in recall.

List Length Experiment 1. First, to eliminate effects of retention interval, we followed previous research and used both the proactive and retroactive conditions, allowing comparisons to be made across blocks equated for study–test lag. Second, to decrease the effect of attention, we interspersed puzzle activity between study blocks. The short lists contained 24 words each, and the long list contained three blocks of 24 words with 3 min of puzzle activity between each block. In this way, participants were required to maintain concentration on each study block for the same amount of time, regardless of whether that block appeared in a short or a long list. In addition, a pleasantness-rating task was used during study. Using a task that requires an explicit response encourages the participant to maintain effort throughout the entire list. Third, the facts that the filler task was much more interesting than the study task and that study was incidental should have discouraged rehearsal. Finally, three different mechanisms were used to decrease the effect of contextual reinstatement. Having the puzzle activity dispersed between the study blocks means that participants were required to reinstate the puzzle context throughout the experiment. This was particularly the case because they were engaged in the same puzzle throughout, so any planning or goal information they may have generated in a previous block could be used to their advantage. To the extent that the puzzle activity formed part of the context during study, it should have served to keep context constant and to facilitate reinstatement. In addition, a distinctive encoding task (pleasantness rating) was used. We anticipated that an encoding task of this nature would focus participants on processing-based forms of context rather than temporal-based forms of context, which are more likely to be affected by length. Furthermore, an 8-min filled retention interval was included at the end of study. If participants used residual temporal context information for a reinstated context vector, then one might expect a context-driven length effect. After 8 min of puzzle activity, however, the residual context information should have deteriorated and should have been much less useful. Under these circumstances, we would expect participants to rely primarily on the processing-based forms of context.

The experimental details are presented in Appendix B. Participants were assigned to one of three conditions: short list at start of study period ($n = 30$), short list at end of study period ($n = 30$), and long ($n = 42$). Participants in the long condition were tested on words from both ends of the list to ensure that they were focusing on the entire list. Consequently, for the long list, the false alarm rates for the start and end of the study period were necessarily identical. Figure 6 shows the hit rates and false alarm rates and their 95% confidence intervals.

When the appropriate controls were in place, there was no significant list length effect, and these data suggest that if the effect did exist, it was very small. In the retroactive design, which has typically shown no effect, our largest difference between either the hit rates or the false alarm rates was 0.025. In the proactive design, which has shown small effects in previous studies, our largest difference was 0.023. The small size of the differences in the proactive design suggests that Underwood's (1978) contention that attention underpins the length effect in this case may well be correct. Furthermore, it seems that by breaking the list into three blocks and including puzzle activity between the blocks, we successfully controlled attention. Note that we also failed to find an effect of lag on recognition performance, suggesting that we were successful as well in focusing participants on the processing components of the context as opposed to the temporal aspects.

Because we used three separated blocks, it is likely that our participants established somewhat different contexts for the three blocks. However, we do not think that this could have helped the participants in the long condition, because the test list contained words from both Block 1 and Block 3. Nevertheless, we decided to eliminate this possibility in a second experiment by presenting all blocks as part of a single list. We were also concerned with the

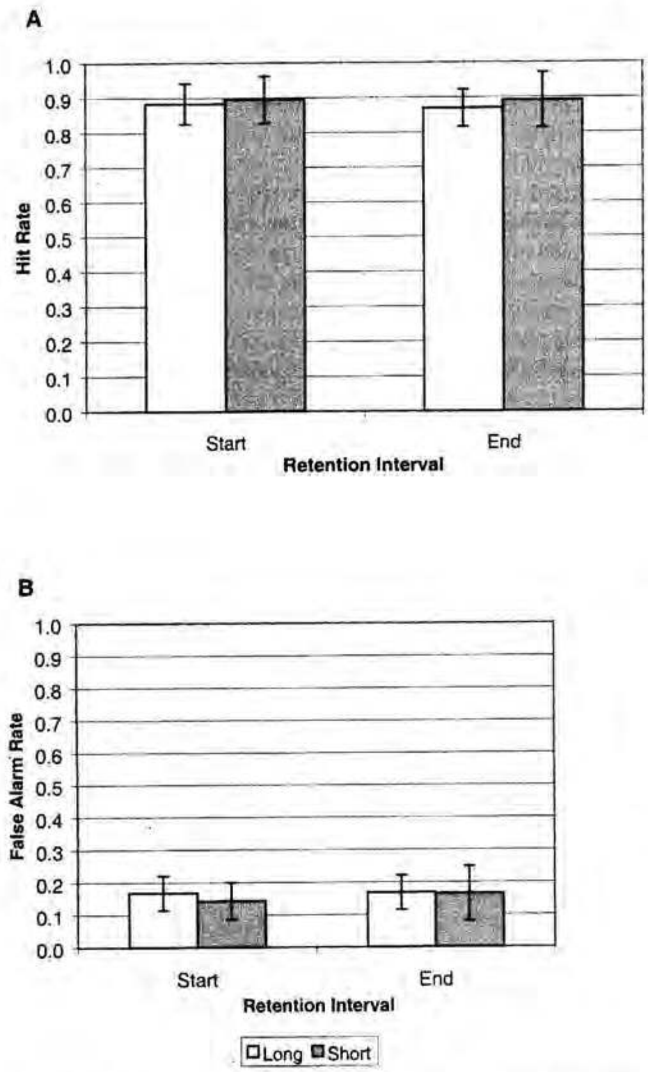


Figure 6. List Length Experiment 1. A: Hit rates. B: False alarm rates. Bars represent 95% confidence intervals.

relatively high level of performance. To reduce the level of performance and to increase the generality of our findings, we decided to use both high-frequency and low-frequency words. In addition, we included a list strength condition because it is the contrast between a null list strength effect and a positive list length effect that has caused so much difficulty for the global matching models.

List Length Experiment 2. In contrast to the previous experiment, a within-subject design was used in which participants took part in all three conditions (short, mixed, and long) during a single experimental session (counterbalanced for order). In the short condition (AB), participants were presented with two blocks of 20 words without a break. In the long condition (ABCD), participants were presented with four blocks of 20 words without a break. In the mixed condition (ABBB), participants were presented with four blocks of 20 words in which the second block was repeated three times. In all cases, the retroactive design was used. Note that this structure provides good control over potential rehearsal because in all three lists participants received two blocks before any differences occurred. However, Block A was the critical test. To influence the results, participants would have had to differentially rehearse Block A items after having seen an intervening block. Puzzle activity followed all conditions to equate study–test lag. Half of the participants received lists of high-frequency words, and half received lists of low-frequency words. Additional experimental details are provided in Appendix C.

Figure 7 shows the results for high- and low-frequency words, including 95% confidence intervals on the critical comparisons. Note that there seems to have been a criterion shift for Block A and new items in the mixed list, most probably as a consequence of the stronger items that appeared in that list. For this reason, we do not make direct comparisons of the hit rates and false alarm rates in the mixed condition. However, we can compare the hits and false alarm rates for the short and long conditions. There were no significant differences in any of the cases. For the high-frequency words, the difference in the hit rates was 0.023, and the difference in the false alarm rates was 0.020. For the low-frequency words, the difference in hit rates was 0.019, and the difference in false alarm rates was 0.028.

Furthermore, if we consider the hit rate minus the false alarm rate (which is equivalent to the interaction term in an analysis of variance on the yes probabilities), we also see small differences. For the high-frequency words the largest of these differences was between the mixed and the long words (0.026), and for the low-frequency words the largest difference was between the long and

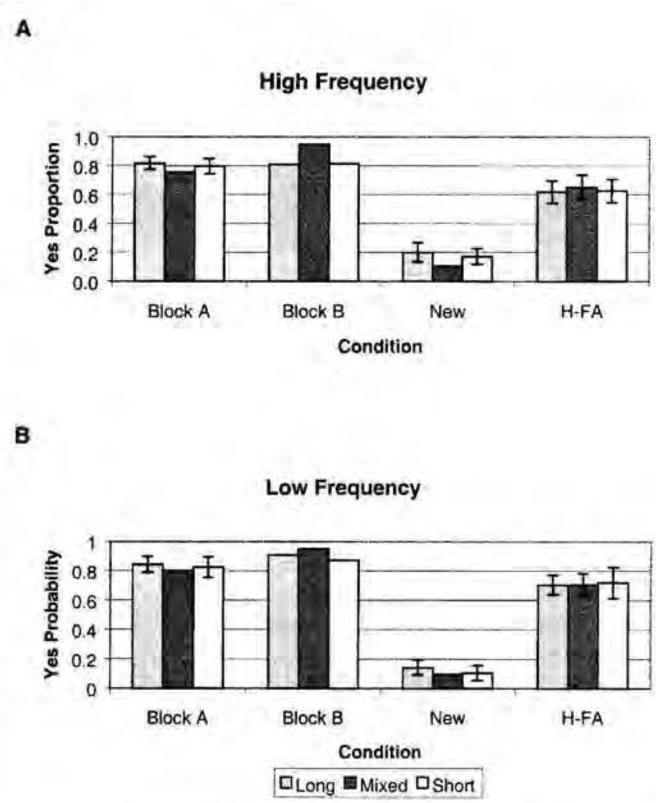


Figure 7. List Length Experiment 2. A: Hit rates, false alarm rates, and hits minus false alarms (H-FA) for high-frequency words. B: Hit rates, false alarm rates, and H-FA for low-frequency words. Bars represent 95% confidence intervals for the comparison of interest (see text).

the short lists (0.009). These differences were nonsignificant. Note, however, that the design was sufficiently powerful to show a significant low-frequency advantage in regard to hits minus false alarms, $F(1, 46) = 5.14$, $MSE = 0.343$, $p = .028$, even though frequency was a between-subjects variable.

Two experiments showed no significant effects of length, and all differences were very small. Our results are consistent with many previous results, but they are inconsistent with others. We have provided a variety of possible explanations for these discrepancies, including lag, inattention, displaced rehearsals, and failures of contextual reinstatement. However, additional research is required to determine which if any of these explanations is correct.

In themselves, the failures to find list length effects do not pose a difficulty for item noise models. A small list length effect can always be accommodated by assuming very low levels of word similarity or high levels of interference from previous lists. However, we expect that item noise models will find it very difficult to simultaneously fit list length effects in recall and recognition. For example, Ohrt and Gronlund (1999) have shown that it is not possible to fit a model such as SAM to recognition and recall list length data without the introduction of a substantial number of additional assumptions. We think that it will be even more difficult with the very small list length effects that we are finding.

In addition, it has been known for some time that cumulative proactive interference paradigms have very different effects on recall-like and recognition-like tasks (Postman & Keppel, 1977). In such designs, participants learn a list and are tested on it 48 hr later. They then learn a second list that is also tested after a 48-hr delay, and so on for subsequent lists. The increasingly large interference effects that are found in recall are simply not revealed in recognition-like tasks such as verbal discrimination.

This differential susceptibility to cumulative proactive interference between recall and recognition is understandable if the interference effects are list length effects. That is, after a retention interval of 48 hr, the reinstated context should be a degraded or noisy version of the study context. This will hurt both recall and recognition, regardless of whether there is a source of proactive interference. However, if there is a source of proactive interference, cuing with the degraded context in recall will activate interfering words as well as target words. However, cuing with the word in recognition will not activate the words in the interfering lists. Thus, our failure to find a list length effect in recognition is compatible with previous failures to find cumulative proactive interference effects in recognition paradigms and is evidence in favor of our assertion that recognition is a context noise process, whereas recall is an item noise process.

Word Frequency Effect

In single-word recognition, words of low normative frequency are recognized better than high-frequency words. Furthermore, performance is worse for both high-frequency targets and high-frequency distractors (the mirror effect; Glanzer & Bowles, 1976). Similarly, if participants are given a forced choice recognition test (i.e., they are presented with two alternatives and asked to indicate which was on the list), they perform better on low-frequency words. In addition, the forced choice design allows the comparison of high- and low-frequency targets and high- and low-frequency distractors. Typically, high-frequency distractors are chosen over low-frequency distractors, and low-frequency targets are chosen over high-frequency targets (Glanzer & Bowles, 1976).

Glanzer, Adams, Iverson, and Kim (1993) have argued that the standard deviations of low-frequency and high-frequency distributions also fall in a specific order. Although it is not possible to observe the standard deviations directly, it is possible to observe the slope of a receiver operating characteristic plotted on z score axes (zROC) to determine the ratio of the standard deviations of two distributions (Green & Swets, 1966). ROC curves are constructed by plotting hit rates against false alarm rates across a number of levels of bias. Typically, bias is altered by varying the probability of the yes response, providing differential rewards, or requiring participants to use a rating scale for the confidence of their judgment (Green & Swets, 1966). Assuming normality of the old and new strength distributions and by taking the z scores of the hit rates and false alarm rates, a zROC curve can be calculated. The intercept of the zROC curve is the d' statistic, and the slope is the ratio of the standard deviations of the new and old distributions. Glanzer et al. (1993) have shown that s(low frequency old/high frequency new) < s(low frequency old/low frequency new) < s(high frequency old/high frequency new) < s(high frequency old/low frequency new), where $s(A/B)$ is the slope of the zROC curve when A is plotted against B. When the assumptions of signal detection theory hold, these results imply that the standard deviation of the high-frequency new distribution is less than that of the low-frequency new distribution and that the standard deviation of the high-frequency old distribution is less than that of the low-frequency old distribution.

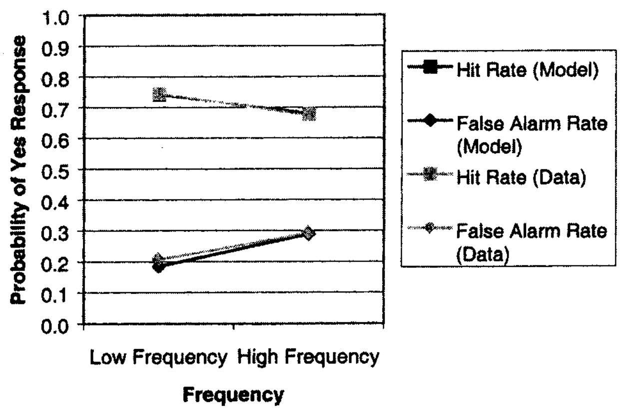
BCDMEM provides a straightforward account of the word frequency data in recognition. Low-frequency words will not be associated with as many contexts as high-frequency words. Therefore, low-frequency words will produce less interference, leading to better performance. Whereas the nature of cuing in BCDMEM compels the low-frequency advantage, it is the decision rule that determines the mirror effect. As outlined in the discussion of the BCDMEM mechanism, as the context noise (p) parameter increases, the distractor distribution approaches one from below and the target distribution approaches one from above.

In this section, we focus on word frequency data from Glanzer and Adams (1990) demonstrating the mirror effect and the ordering of zROC slopes.[2] In a series of five experiments, participants were given a study list on which they performed lexical decision. Words were presented either auditorially or visually under either intentional or incidental learning conditions. The study list was followed by a recognition test. For each test word, participants indicated whether they thought it was old and then made a confidence rating.

Although there was substantial variability depending on the learning conditions, the pattern of results for both the probability of a yes response (i.e., hits and false alarms) and the slope ratio was consistent across all five experiments. For simulation purposes, we have chosen to model the averaged data (see Figure 8).

[2] Dennis and Humphreys (1998) also reported simulations of word frequency. The results reported here, however, are based on the average data across five experiments rather than on a single experiment and have been extended to consider the ordering of zROC slopes.

A



B

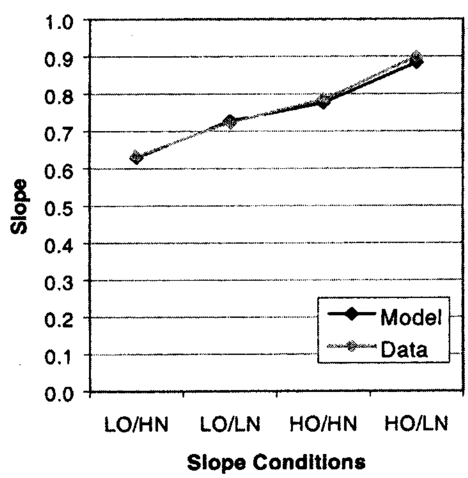


Figure 8. Fit of the bind cue decide model of episodic memory to word frequency data of Glanzer and Adams (1990). A: Fit to hits and false alarm rates. B: Fit to the *z*ROC (receiver operating characteristic plotted on *z* score axes) slope data. LO = low frequency old; HN = high frequency new; LN = low frequency new; HO = high frequency old.

In simulating the data, a single parameter was estimated for amount of learning (r) and degree of contextual reinstatement (d). The context noise parameter (p) was varied for the low-frequency and high-frequency words. Context noise (p) incorporates the number of other contexts in which the word has been seen and the amount of learning in those contexts. As a first approximation, we expect it to increase with word frequency. However, including additional contexts is likely to affect the context noise more than repeating a word within the same context (because the greater the number of contexts, the more likely it is that one of them will resemble the current context). Therefore, the number of different contexts in which a word appears may be a more sensitive measure of context noise (because the context representations are sparse and chosen independently). Because the correlation between the number of contexts in which a word has appeared (operationalized as the number of articles in which it appeared in the 1994 editions of the *Sydney Morning Herald;* Dennis, 1995) and its frequency is very high (r^2 = .95; Dennis, 1995), we assume that word frequency reflects context noise under most conditions. In addition, five criteria were used to allow the estimation of *z*ROC slopes. One was set at 1.0, which is the normative figure, whereas the other four were optimized. Figure 8A shows the model's fit of the hit and false alarm rates plotted on the same graph as the original data (using the normative criterion). The parameters of the fit were as follows: learning rate, r = .602; contextual reinstatement, d = 0.590; and context noise, p(low) = .094 and p(high) = .336. Figure 8B shows the fit for the *z*ROC slopes (the criteria were 0.007, 0.556, 1.0, 2.975, and 3.965).

The fit for both the hit and false alarm rates and the *z*ROC slopes was good. For the hit and false alarm rates, the maximum absolute difference was 0.002 and the data-to-model correlation was .999. For the *z*ROC slopes, the maximum absolute difference was 0.015, and the model-to-data correlation was .999. BCDMEM is capable of accounting for the mirror effect for word frequency and the slopes of the *z*ROC distributions.

Reder et al. (in press) have also proposed that the larger number of previous contexts associated with high-frequency words is responsible for the lower hit rate with these words. This is the case because activation is spread from the item to the contexts associated with that item in proportion to the total number of associated contexts. However, this model also spreads activation from the context node to the list items, so it will predict a list length effect and possibly a list strength effect.

Concreteness is another word characteristic that has been shown to mirror, and we would like to suggest that the basis of this effect is the same as that for the word frequency effect. Participants may be better able to recognize concrete words than abstract words (Glanzer & Adams, 1990) because concrete words appear in fewer contexts. We found that the number of contexts in which a word had appeared (taken from the *Sydney Morning Herald* database; Dennis, 1995) was correlated negatively (−.26) with concreteness ratings (taken from the Medical Research Council Psycholinguistic Database [Coltheart, 1981]). Furthermore, when word frequency and concreteness were used to predict number of contexts in a regression analysis, there was a significant advantage to adding concreteness ($p < .0001$). Consequently, even when word frequency is controlled (as in Glanzer & Adams, 1990), BCDMEM will still predict a concrete advantage.

Similarity Effects

Item noise models of recognition predict that participants will be more likely to produce a false alarm to a lure that is similar to one of the study words. In item noise models, overlap between the test word and a study word produces a strong match, making it more likely that the global match will exceed the criterion. A number of studies have demonstrated substantial "false memory" effects (Brainerd, Reyna, & Mojardin, 1999; Deese, 1959; Postman, 1951; Roediger & McDermott, 1995; Shiffrin, Huber, & Marinelli, 1995). By contrast, in other studies the effect has been substantially smaller and somewhat unreliable (Anisfeld & Knapp, 1968; Grossman & Eagle, 1970; MacLeod & Nelson, 1976; Mandler, Pearlstone, & Koopmans, 1969; Underwood & Humphreys, 1979). This wide variation suggests that something more than sheer similarity is involved.

One possibility is that participants implicitly produce associates of the study words (implicit associative responses). At test, when the associate is presented, it is "as if" it had appeared on the list, and hence the participant is more likely to say yes. Such a mechanism is more probable in paradigms in which the related words are presented in a blocked fashion so that participants are aware of the categorical nature of the list and can encode the nature of that relationship (e.g., they may implicitly generate category labels or other category members). Deese (1959) and Roediger and McDermott (1995) used such paradigms, and it is these paradigms that typically show large false memory effects.

In addition, Anisfield and Knapp (1968) reported that the false recognition effect was directional. That is, if *A* elicits *B* in free association but *B* does not elicit *A*, studying *A* increases the probability of falsely recognizing *B*, but studying *B* does not increase the probability of falsely recognizing *A*. This directionality is difficult to reconcile with a similarity mechanism as used in item noise models.

An alternative explanation of the false memory effect is that participants may adjust their criteria for accepting members of a category that they know appeared on the list. Miller and Wolford (1999) assessed this possibility using lists similar to those of Roediger and McDermott (1995) that included some prototype items. They found that there was a criterion shift but no impact on d' values, a result inconsistent with an item noise account.

Shiffrin et al. (1995) recognized the possibility that study trial encoding or a criterion shift could be responsible for certain false recognition effects and took steps to reduce the possibility that their participants would become aware of the categories used in the study list. They used materials similar to those used by Deese (1959) and Roediger and McDermott (1995) but embedded them in a long list in which the instances of any given category were widely spaced. In this design, false alarms to nonpresented category instances and prototypes (the word associated with each member of the category) also increased substantially with the number of category exemplars studied. However, Shiffrin et al. (1995) did not include a control condition to determine whether the categorical nature of the lists influenced the false alarm rate of related lures.

Dewhurst and Anderson (1999), however, did use such a control. They had participants study lists containing categories of one, four, and eight words distributed throughout the study list. In the control condition, instead of different category exemplars being presented, the same item was repeated multiple times (one, four, or eight). Consequently, in the control condition, there was no categorical structure for participants to extract. For the related lures for categories of length one, item noise models predict no difference as a function of the categorical structure of the remainder of the list. However, Dewhurst and Anderson (1999) found an increase in false alarm rate in the categorized list despite an overall criterion shift in the opposite direction. This result provides strong evidence that the categorical nature of the list (even in lists in which categories are distributed throughout) induces participants to respond on the basis of category membership in a way that is inconsistent with an item noise approach.

Dewhurst and Anderson's (1999) study was not designed to examine the impact of the categorical nature of the list and, as a consequence, did not counterbalance the related and unrelated lures. Similarly, other studies have compared false alarms to words that have a special characteristic (e.g., they are synonyms) and control words that do not have that special characteristic (e.g., Brainerd et al., 1999). In these designs, no matter how carefully the control words have been matched on relevant dimensions (e.g., frequency), there is always the possibility that word differences are inflating the false alarm rate of the experimental words and generating an artifactual false memory effect.

The effects of word similarity on false alarm rates clearly pose a problem for BCDMEM and other context noise models. However, most of the time when there is a single synonym, antonym, category label, or associate of a recognition lure in the target list, the effects are quite small. On those occasions when a large effect has been found, it may be attributable to implicit associative responses or category-based criterion shifts or to the use of imperfectly matched control lures. Further research, including research on discriminating between internally generated words and list words as a function of study modality, is required.

BCDMEM and the Dual-Processing Approach

As outlined earlier, the dual-processing approach proposes that episodic recognition involves two independent processes. The familiarity process is context insensitive and automatic, whereas the recollection process is context sensitive and strategic. To assess the contributions of familiarity and recollection to the recognition decision, Jacoby (1991) introduced the process dissociation procedure. In a typical application of the process dissociation procedure, participants study two lists and are asked to make one of two recognition decisions at test (Jacoby, 1991). In the *inclusion* condition, they are required to respond yes if the word was present in either of the two lists. In the *exclusion* condition, they are required to respond yes only if the word appeared in the target list (either List 1 or List 2). Yonelinas (1994) had participants perform just the second task (i.e., saying yes only if the word appeared in the target list). The inclusion probability was defined as the probability that a participant said yes to a word from the target list. The exclusion probability was defined as the probability that a participant said yes to a word from the nontarget list. Under the dual-processing logic, the nonstrategic familiarity process is assumed to lead to a yes response in both the inclusion and exclusion conditions, whereas the strategic recollection process is assumed to allow the rejection of words in the exclusion condition. When the process dissociation procedure is used, estimates of the contributions of familiarity and recollection can be made (Jacoby, 1991; Yonelinas, 1994).

For our purposes, there are three important attributes of the dual-processing approach. First, the dual-processing approach has emphasized the way in which a participant has been instructed to process the words within a list as a source of discriminating information and not purely as a determinant of the strength or depth of processing (see also the source monitoring framework of Johnson, Hashtroudi, & Lindsay, 1993). For instance, Jacoby (1991, Experiment 3) required participants to either read or form anagrams with the words from the first list and listen to the words from the second list. As a consequence, any information that was retained about the nature of the task that a participant was required to complete with the word could be used to distinguish the list in which it appeared. Within BCDMEM, we call this information "processing context" (see the Context and Its Reinstatement section) and assume that it is associated with the word at study in the same way in which other forms of context are stored.

Second, within multilist paradigms, in which a participant knows that if a word was present in one list, it was not present in another list, successful recollection can allow the participant to exclude or discount a word. For instance, in Jacoby's (1991) design, a strong match to an "anagram" reinstated context vector could be used to reject the word. Because the second list did not involve solving anagrams and participants are unlikely to have solved an anagram for a given word before the experiment, the match between the retrieved context to a word that had been solved as an anagram and a reinstated anagram context would provide good evidence that the word appeared in the first list. Discounting (though not necessarily through recollection) is necessary for understanding multilist paradigms, and, in the application of BCDMEM to Jacoby's (1991) data, we assume that participants match against an experiment-wide context to eliminate new words and then an anagram–read context to eliminate List 1 words.

Third, the dual-processing approach has emphasized the use of a retrieved context in making a decision as to whether the to-be-recognized item occurred in the experiment or in a particular list (see also Johnson et al., 1993). This is common ground with context noise models, although the dual-processing approach and the source monitoring framework use the retrieved context in a symbolic inference process, whereas BCDMEM uses a subsymbolic inference process. Furthermore, neither the dual-processing approach nor the source monitoring framework is committed to the hypothesis that the previous contexts in which a word has occurred are significant sources of noise.

We now turn our attention to the modeling of process dissociation data using BCDMEM. Ratcliff et al. (1995) have demonstrated that the data of Yonelinas (1994) in which list length was manipulated require only a single-process model. We begin by showing how BCDMEM accounts for these data using the decision rule alone and then demonstrate that the same is true of data on temporal separation (Hall, 1996). When strength and type of study task are manipulated, it becomes necessary to use discounting. In the subsequent section, we first report data in which strength was manipulated and show how BCDMEM can account for the results through discounting. Then we model data from Jacoby (1991) on the effect of manipulating the study task.

List Length and the Process Dissociation Procedure

Yonelinas (1994, Experiment 1) presented data showing the effect of list length (confounded with study–test lag). Each participant completed six sessions and each session involved eight study–test blocks. Each block contained two study lists followed by two test lists. The study lists were either both short (10 words) or both long (30 words). In the first recognition test, participants were to respond yes to List 1 words only. In the second recognition test, they were to respond yes to List 2 words only. Each test list contained List 1, List 2, and new words. Inclusion words were those for which the target list at test was the list in which they were presented. Exclusion words were those for which the target list at test was the opposite from that in which they were presented. Figure 9 shows the data obtained.

To apply BCDMEM to these results, we could assume that participants use an end of list context when they are asked to say yes to List 2 words and a reinstated List 1 context when they are asked to say yes to List 1 words. The alternative is to assume that both the List 1 and List 2 contexts are reinstated. The data provided did not allow us to distinguish between these two alternatives (i.e., separate results were not provided for List 1 and List 2). In addition, Yonelinas's participants knew that they had to differentiate between lists, and they were well practiced. We thus assume that his participants established separate representations for the List 1 and List 2 contexts and that, on the test, they reinstated the appropriate context. This is essentially the assumption that Yonelinas (1994) made about why his participants could perform the task in the absence of a differential processing task. We included an additional parameter representing the overlap (o) between the List 1 and List 2 study contexts. The overlap parameter represents the probability that a given component is a one in both study contexts. If the two contexts were being formed independently, the overlap parameter would equal the square of the sparsity parameter ($.02^2 = .0004$). Obtained values of the overlap parameter that are lager than this value indicate that the contexts for the two lists are more similar than would be expected by chance either because they occur close to each other in time or because they describe similar processing operations. This parameter was used in the construction of the contexts for List 1 and List 2.[3]

Figure 9 shows the fit of the model to the data. The parameters of the fit were as follows: learning rate, $r = .259$; context noise, $p = .089$; context overlap, $o = .003$; and contextual reinstatement, $d(\text{long}) = 0.608$ and $d(\text{short}) = 0.337$. The maximum absolute difference was 0.025, and the correlation was .99.

Temporal Separation and the Process Dissociation Procedure

In Yonelinas's (1994) experiment reported in the previous section, the participants were well practiced. Under these conditions, it is conceivable that the list context was essentially a label (J. A. Anderson & Bower, 1974). To examine the effect of temporal context in a situation in which participants were less likely to label the two lists, Hall (1996) varied the temporal separation between the lists among unpracticed participants. In the after lists condition, participants studied List 1 and then List 2 and spent 8 min solving a puzzle task before being tested. In the between lists condition, the sequence was List 1, puzzle, task List 2, test. Jacoby's (1991) version of the process dissociation procedure was used with inclusion instructions that covered both lists and exclusion instructions that targeted List 2. The critical words were those from List 1, and the design

[3] To construct context vectors of sparsity s that overlapped to different degrees, C1 was chosen first by setting each component to one with probability s. Next, the components of C1 that would be involved in the overlap were chosen from the components of C1 that were set to one with probability o/s (where s is sparsity). To complete C2, those components that were not equal to one in C1 were set to one with probability $(s - o)/(1 - s)$.

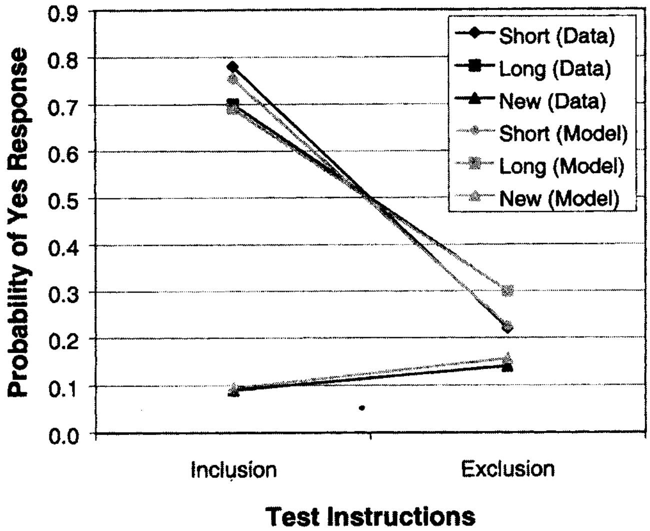


Figure 9. Fit of the bind cue decide model of episodic memory to data of Yonelinas (1994, Experiment 1).

controlled for the recency of these words. In both study lists, participants were asked to make pleasantness ratings.

Figure 10 shows the results. Whereas the inclusion results for the List 1 words differed very little as a function of List 2 placement (.825 vs. .865), the exclusion probability of the List 1 words was much lower when the lists were separated by the filled interval than when they followed each other (.365 vs. .590).

To model the manipulation of interlist interval in BCDMEM, we allowed the overlap parameter to change. Recall from the previous section that the overlap parameter is the probability that a component is a one in the context vectors of both lists in the process dissociation paradigm. Placing the filled 8-min interval between the lists should lead to a lower value of this parameter (i.e., a decrease in the similarity of the List 1 and List 2 context vectors).

Bias may also have played a role in Hall's (1996) results. All of the exclusion probabilities were below the corresponding inclusion probabilities, suggesting the use of a more stringent bias in exclusion (cf. Buchner et al., 1995). In the exposition of the BCDMEM likelihood ratio, we argued that prior odds could be eliminated on the basis that, in most experiments, the probability of an old word is equal to the probability of a new word. The probability that a word will invoke a yes response is higher under inclusion instructions than in exclusion. In Hall's (1996) experiment, the actual probabilities of words to which participants should respond yes were .66 in the inclusion case and .33 in the exclusion case. Although it is unclear how accurate participants might be in regard to estimating prior odds, the results suggest that these odds play a role. Rather than add two new free parameters to model the prior odds, we chose to set the exclusion probability to .33 and allow the inclusion probability to be optimized. Figure 10 shows the fits to the data for the between lists interval (Panel A) and the after lists interval (Panel B). The parameters of the fit were as follows: inclusion prior, .909; learning rate, r = .354; context noise, p = .160; context overlap, o(between) = .001 and o(after) = .015; and contextual reinstatement, d = 0.337. The maximum absolute difference was 0.056, and the correlation between the model and the data was .993. Therefore, BCDMEM seems to have captured the effect of temporal separation.

One caveat is that the estimated value of the inclusion prior parameter (.909) appears to be high considering that the prior odds in the inclusion condition were .66. However, a lower value could have been obtained if we had allowed both priors to vary, and it is unclear how accurately the participants may have been able to estimate the priors given that they were exposed to only a single test list.

Strength and the Process Dissociation Procedure

Strength, as measured by either number of repetitions or study time, is an important variable to distinguish between process accounts of data obtained with the process dissociation procedure. SAM (and the other global matching models) predicts that increasing strength on the words from the to-be-excluded list will make them less likely to be excluded (Mulligan & Hirshman, 1997;

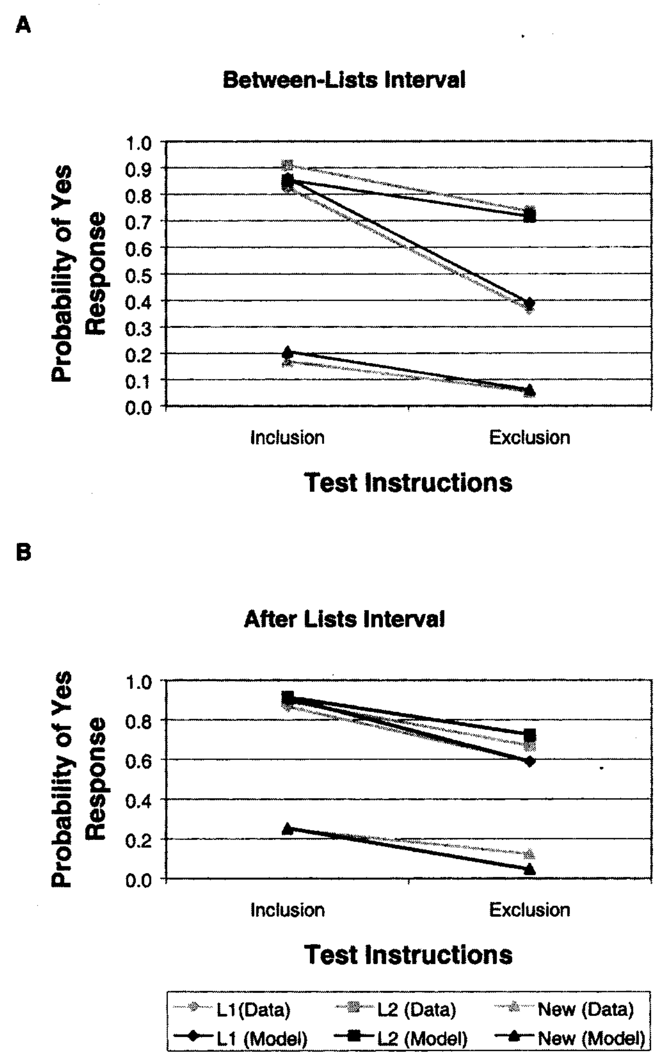


Figure 10. Fit of the bind cue decide model of episodic memory data of Hall (1996, Experiment 1). A: Fit for the between lists interval. B: Fit for the after lists interval. L = list.

Ratcliff et al., 1995), provided it is not possible to manipulate the criterion between the strength conditions. In Jacoby's (1991) Experiment 3, however, the strong words (i.e., anagram words) were excluded more successfully than the weak words (i.e., read words). For this reason, the single-process global matching accounts are not sufficient (Ratcliff et al., 1995), although they could be augmented with a discounting rule to account for the data.

However, in the experiments that have been conducted to date, strength has been manipulated by a change in study task. It may be, however, that manipulating the study task (i.e., anagram solving vs. reading) introduces both qualitative and quantitative changes in the memory of the word. For instance, the memory for a word for which the participant solved an anagram may be both stronger and more unique than that for a read word.

In the current experiment (see Appendix D for experimental details), strength was manipulated by presenting words either one or four times to determine the effect independent of a study task difference. In addition, Yonelinas's (1994) process dissociation design was used, in which either the first or the second list was identified as the to-be-recognized list. Inclusion words were those from the to-be-recognized list, whereas exclusion words were those from the other list.

The result of prime interest is whether participants are able to use List 1 and List 2 context cues with equal efficiency. Figure 11 shows the probability of a yes response in the exclusion condition as a function of presentation list and number of presentations. When a word is presented in List 1 (i.e., we ask participants to say yes if they believe that it appeared in List 2), the probability of a yes response rises with the number of presentations. Additional strength makes it more difficult to correctly exclude the word. In contrast, when a word is presented in List 2, the probability of incorrectly saying yes in the exclusion condition decreases as a function of number of presentations. The bars in Figure 11 are the 95% confidence intervals for these two within-subject comparisons considered separately.

These results suggest that participants use an experiment-wide context cue and a List 2 context cue but not a List 1 context cue. Repeating a word increases the strength of the experiment context association, increasing the probability of a yes response. When List 1 is the target list, the List 2 context cue is used to try to eliminate List 2 intrusions. Consequently, the probability of saying yes to a List 2 word decreases with strength. However, the same process does not seem to occur for List 1 words. Participants may find it difficult to reinstate the List 1 context when there is no difference in the study task between List 1 and List 2.

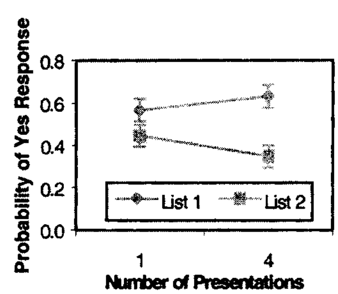


Figure 11. Process dissociation strength data: Probability of a yes response with exclusion instructions as a function of presentation list and number of presentations.

To simulate these results in BCDMEM, it was assumed that participants compare the retrieved vector against an experiment-wide reinstated context vector. This vector was taken to be the bitwise odds ratio of the context vectors representing List 1 and List 2. The model responds no if the likelihood ratio is below the criterion eliminating new words. For words that are above the criterion (i.e., words that the system believes did occur in one of the lists), a second comparison is made against the reinstated List 2 context vector. When the target list is List 2, an above criterion likelihood ratio leads to a yes response. If the target list is List 1, discounting is used. An above criterion likelihood ratio leads to a no response. Figure 12 shows the fit of the model to the data. The parameters for this fit were as follows: experiment context prior probability, .770; Target List 1 prior probability, .046; Target List 2 prior probability, .767; learning rates, r(strong) = .686 and r(weak) = .555; context noise, p = .452; context overlap, o = .009; and contextual reinstatement, d = 0.542. The maximum absolute difference was 0.046, and the correlation of the data with the model was .993. The model shows the interaction of strength and target list outlined earlier.

Study Task Discriminability and the Process Dissociation Procedure

In the preceding three sections, the nature of the study task was the same in both lists. The differences between the List 1 and List 2 context vectors under these conditions are expected to be a consequence of the passage of time and words rather than the nature of the processing task. In Jacoby's (1991) original experiment, however, study task was manipulated between lists. The first list contained words that were solved as anagrams and words that were read, whereas the second list involved words that were heard. Figure 13 shows the results for Jacoby (1991).

There is an important difference between the strength results reported in the previous section and the Jacoby case. In the previous section, the crossover pattern was seen only when the to-be-recognized list was List 1. It did not occur when the to-be-recognized list was List 2. This was explained by assuming that participants were unable to reinstate a List 1 context. In the Jacoby study, however, the to-be-recognized list was always List 2, yet the crossover pattern was obtained. The difference may reside in the nature of the study tasks. In Jacoby's design, the first list involved an anagram context, which is likely to be distinctive and easy to reinstate.

In modeling Jacoby's data, we assumed (as in the preceding section) that participants start by making an experiment-wide judgment about whether the word appeared in either list (using an experiment-wide context vector) in both inclusion and exclusion. In exclusion, however, we assumed that participants can and do reinstate a List 1 context, using this to exclude words from List 1 through discounting.

Figure 13 shows that, by applying these assumptions, we can fit the data. The parameters of the fit were as follows: list-wide prior, .557; List 1 prior, .073; learning rates, r(anagram), .983, r(read), .416, r(heard), .775; context noise, p = .365; context overlap, o = .0100; and contextual reinstatement, d = 0.832. The maximum

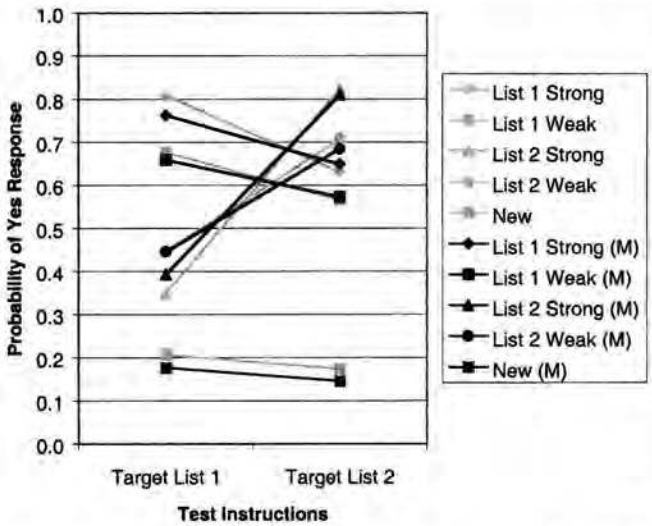


Figure 12. Fit of the bind cue decide model of episodic memory to process dissociation strength data outlined in Figure 11. M = model.

absolute difference was 0.045, and the correlation of the data and the model fit was .992.

In summary, BCDMEM is able to capture data from the process dissociation procedure (e.g., effects of length and type of processing). The key theoretical contribution is the emphasis on the reinstatement of context. BCDMEM argues that participants have substantial control over the nature of the context(s) against which they compare the retrieved context vector and that the context vector that they reinstate depends on the circumstances. When study tasks distinguish the study lists from previous lists and the study lists from each other, participants will reinstate a context vector based on this information. When the study task is not distinctive, participants revert to temporal context (and possibly only the most recent context). They may be capable of organizing contextual representations hierarchically, forming experiment-wide and list-wide contexts, and they seem capable of using matching information to discount words that appeared in other contexts. Understanding the ability to reconstruct context and the strategies by which it is used will be an important objective for further research, and the process dissociation procedure provides an empirical mechanism with which to proceed. In the next section, we outline two studies that demonstrate the productivity of the approach.

Multiple List Designs With Manipulations of Processing and Temporal Similarity

The BCDMEM decision rule involves comparing the retrieved context vector representing the contexts in which a test word has appeared against a reinstated context vector representing the contexts for which the participant is looking. Multilist designs allow the manipulation of both of these representations and hence provide critical tests of the context noise approach.

BCDMEM predicts that performance will decline if the diffuseness of either the reinstated or the retrieved context increases. For instance, if a word appears in multiple contexts that overlap with the reinstated context, it should be difficult to exclude. To test this prediction, we used a three-list paradigm (see Appendix E for additional experimental details). The lists were presented in immediate succession, and we manipulated the processing requirements so that List 1 and List 3 had a processing component in common, and List 2 and List 3 had another processing component in common. The words in List 1 were read and rated for pleasantness, the words in List 2 were heard and typed (the typed letters were not displayed on the screen), and the words in List 3 were heard and rated for pleasantness. Here participants were asked to say yes to List 3 and no to List 1, List 2, and new words. They were correctly informed that if a word appeared in either List 1 or List 2, it did not appear in List 3. The instructions also reminded the participants as to how the words in Lists 1, 2, and 3 had been processed. The critical manipulation involved the presentations of words in List 1 and List 2. Words appeared once in List 1, once in List 2, twice in List 1, twice in List 2, or once in List 1 and once in List 2. The retrieved context for words that occurred once in List 1 and once in List 2 should be more diffuse than for words in the other conditions. That is, for these words, the retrieved context would contain the shared component between List 1 and List 3, the shared component between List 2 and List 3, and the unique components in List 1 and List 2. In contrast, a word presented exclusively in List 1 or exclusively in List 2 would contain only the shared component between that list and List 3 along with the unique component. We present the results from two experiments that differed only in the post–List 3 retention interval. We also included an inclusion condition in which participants were asked to say yes to List 1, List 2, and List 3 words and no to new words. Because the pattern of results was basically the same in the two experiments, we report the combined results.

In Figure 14, we provide the probability of yes responses to new words, the average of the words presented once in List 1 or once in List 2, the average of the words presented twice in List 1 or twice in List 2, words presented once in List 1 and once in List 2, and words presented once in List 3. We also provide the 95% confidence intervals for the words presented once, twice, and once in both Lists 1 and 2.

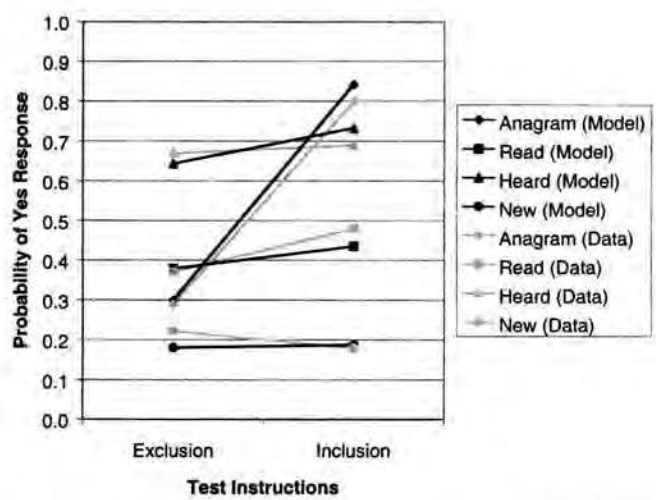


Figure 13. Fit of the bind cue decide model of episodic memory to data of Jacoby (1991, Experiment 3).

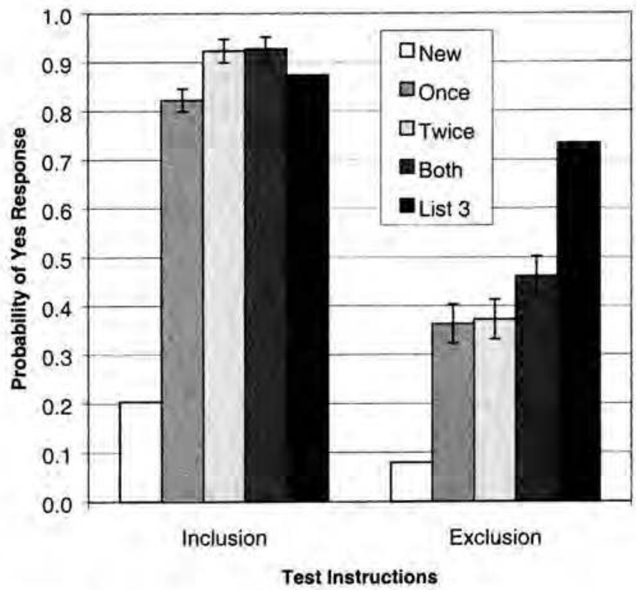


Figure 14. Diffuseness of retrieved context. Bars represent 95% confidence intervals for the comparisons of interest (see text).

The results for the inclusion conditions conform to what we would expect to be everyone's a priori predictions. Participants were capable of discriminating between old and new words, and the probability of saying yes to repeated words was higher than the probability of saying yes to nonrepeated words. In addition, the probability of saying yes to once-presented List 3 words fell between the probabilities of saying yes to once-presented and twice-presented words from the earlier lists.

The results for the exclusion conditions probably violate some a priori expectations. They were, however, in accordance with the predictions of BCDMEM. When a word was repeated twice in the same to-be-excluded list, it did not increase the probability that it would be falsely identified as a List 3 word. This may have occurred because within-list repetitions enhanced the ability to discount or reject words from a to-be-excluded list. In contrast, the repetition of a word across two different lists did increase the probability that it would be falsely identified as a List 3 word. So the fact that participants were looking for a heard word that was rated for pleasantness increased the false alarm rate for the words that appeared in the first two lists because they were both heard and rated for pleasantness, albeit in different lists (as BCDMEM predicts).

The preceding design involved manipulating the content of the retrieved context vector. Multilist designs also allow the manipulation of what the participant is looking for, that is, the reinstated context vector. In an unpublished study conducted by Kerry Chalmers, the diffuseness of the reinstated context was manipulated. Participants first completed a familiarization session on a set of very-low-frequency words. In the familiarization session, words were presented with their definitions six times (see Chalmers & Humphreys, 1998, for details about a set of very similar experiments). One week after the familiarization session, participants returned for a series of study–test sessions. In each case, half of the study words had been previously familiarized, and half were unfamiliarized. Similarly, half of the distractors on the test had been familiarized, and half were unfamiliarized. In the immediate test condition, the List 2 and List 3 words were presented as a single study list, and the test immediately followed the presentation of the last List 3 word. In the delayed test condition, the List 2 and List 3 words were also presented as a single study list, but the test was delayed for 24 hr. In the mixed condition, after studying the List 2 words, the participants were asked to return the next day. When they returned, they studied the List 3 words and were then given the test for both the List 2 and List 3 words. The test instructions informed the participants that they should say yes to words present at study regardless of whether or not they had also been present during the familiarization phase. Participants were told that words could occur both during familiarization and at study.

In interpreting these results, we assume that participants in the immediate condition can either reinstate a relatively precise study context or simply use the end-of-list context. In either case, we expect that the context they use or reinstate will not overlap much with the context of the familiarization phase. The result will be a relatively low false alarm rate for the words not present at study but present at familiarization. In the delayed condition, it will be necessary to reinstate the previous day's study context. This reinstatement should be relatively difficult. As a result, we expect that the context in the delayed condition should have more overlap with the context of the familiarization phase, and there should be an increase in the false alarm rate. This prediction, however, is relatively uninteresting. A more interesting prediction concerns the false alarm rates in the mixed and delayed conditions. Many models predict that the false alarm rate in the mixed condition will fall between those for the immediate and the delayed conditions. In contrast, BCDMEM assumes that the reinstated context in the mixed condition is more diffuse than the contexts in the other two conditions because it must specify words that appeared on 2 different days. Consequently, BCDMEM predicts that the false alarm rate will be highest in the mixed condition.

These predictions were supported. The hit rates and false alarm rates, along with the 95% confidence intervals for the familiarized words are shown in Figure 15. Most important, the false alarm rate for the

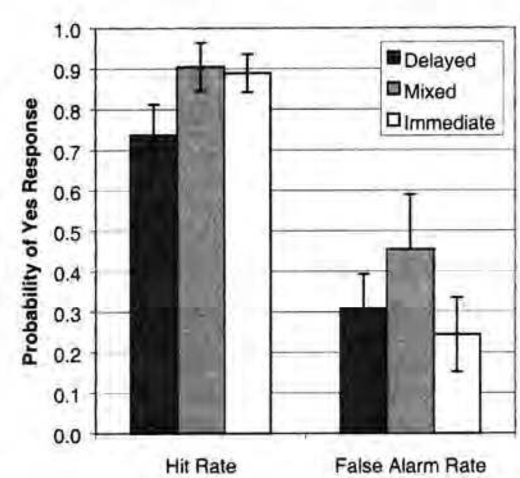


Figure 15. Diffuseness of reinstated context (results from an unpublished study conducted by Kerry Chalmers). Bars represent 95% confidence intervals for the comparison of interest (see text).

mixed condition was the highest of the three and certainly higher than the midpoint of the immediate and the delayed conditions.

General Discussion

The three approaches we have considered—item noise, dual processing, and context noise—each have their natural domains of application. The domain of the item noise models is the single-list paradigm in which list length, list strength, or word similarity is manipulated. The domain of the dual-processing approach is the two-list inclusion–exclusion paradigm with the introduction of a study or test manipulation such as time pressure. BCDMEM's domain is the multiple-list paradigm in which temporal or processing similarity is manipulated. Within their respective domains, all three approaches provide reasonably parsimonious explanations. Outside their domains, the explanations can be considerably less parsimonious.

BCDMEM is no exception. It provides relatively parsimonious explanations for the effects of word frequency on recognition, the effects of manipulating the diffuseness of the reinstated and retrieved contexts, the null list strength effect, the finding that in some situations there is also a null list length effect, the lack of cumulative proactive interference in recognition, and the thresholdlike behavior involved in recognizing morphemic relatives. It also provides an explanation, in common with the other Bayesian models, for the ubiquity of the mirror effect. However, although we think that similarity effects and the significant list length effects that have been reported in the literature can be explained by a variety of processes, BCDMEM clearly does not provide a parsimonious explanation for these findings. In addition, the application of BCDMEM to the process dissociation procedure required a variety of post hoc decisions about what contexts were being reinstated.

At this time, however, BCDMEM is the only model to clearly address all three domains. Furthermore, the number of parameters has been determined primarily by the complexity of the domains and is not excessive in comparison with other models. Direct comparisons of the number of parameters across models are somewhat difficult, because different parameters tend to play more or less central roles in each model. BCDMEM has five main parameters: vector length, sparsity, learning rate, context noise, and contextual reinstatement. Of these, only learning rate, context noise, and contextual reinstatement are optimized in simulations. In addition, a prior probability parameter is added when participants manipulate the criterion, and a contextual overlap parameter is added for dual-list paradigms. In contrast, REM.5 has three main parameters: vector length, learning rate, and the parameter governing feature generation. In addition, there are parameters for the number of storage attempts when words are superimposed, the ratio of features in the context and the word, the activation threshold over which a word is included in the activated set, and the rate of contextual drift. Furthermore, parameters will be required when the criterion is manipulated through instructions and when dual-list paradigms are considered. It would seem, then, that BCDMEM is at least as economical in regard to parameters as item noise alternatives. A more meaningful criterion, however, is how tightly model parameters are tied to experimental manipulations. BCDMEM does well in this regard. Only vector length and sparsity are not directly related to experimental variables, and neither of these parameters is optimized.

BCDMEM makes commitments on five crucial issues that contrast it against existing models of episodic recognition: (a) the content of context, (b) the role of discounting or dual processes in recognition, (c) the use of a subsymbolic inference process, (d) the use of a decision rule that approximates a likelihood calculation, and (e) the emphasis on context noise. Each of these issues is addressed in the following discussion.

Content of Context

BCDMEM postulates that there are at least two forms of context, processing and temporal. This emphasis on two forms appears to be relatively unique and stems from the fact that the model is designed to account for data from both the item noise and dual-processing domains. The item noise approach, with its interest in single-list designs and the other items in the list, has emphasized either an abstract form of context or a temporal context (Gillund & Shiffrin, 1984; Ratcliff et al., 1995). To extend the model to the dual-processing domain in which words are commonly processed in different ways within the same study–test cycle, it was necessary to expand the idea of context to include information about the task set for the participant.

Postulating two types of context does not, by itself, increase the explanatory power of a theory. That is, we already know that participants can discriminate between two lists. Postulating that this ability results from a hypothetical entity (context) that changes between List 1 and List 2 does not add to the basic observation. It is only when the postulated entity enters into new predictions and other explanations that a real advance has been made. We believe that BCDMEM fares well in this regard. Our examination of the number of contexts in which words occur and manipulations of the diffuseness of retrieved and reinstated contexts are steps toward providing the concept of context with genuine explanatory power.

Discounting or Dual Processes

There are no viable single-factor models of recognition, and BCDMEM is no exception. In fitting the data from the process dissociation procedure, we reinstated two different contexts. Furthermore, at times we used discounting to reject words that appeared in other lists. Thus, the information being used to reject the items from one of the lists was different than the information being used to accept the items from the other list. Within the BCDMEM framework, two types of information, but not two fundamentally different retrieval processes, are used (see the next section).

Subsymbolic Inference

The distinction between the symbolic and subsymbolic levels of description becomes important in evaluating the alternative decision mechanisms that can be used when we cue with the word for contexts in which that word has appeared. In BCDMEM, the retrieved vector is a composite of the context vectors associated with the word. With the decision mechanism we use, there is no need to resolve the conflict between the competing associations before the retrieved information can drive the inferential process.

That is, the retrieved vector has no meaning in its own right; it is meaningful only when it is compared with the reinstated vector. The alternative is that the competition is resolved or the noise is eliminated so that the retrieval process can produce symbolic information such as a concept, a word, or a significant feature (Clark, 1999). The retrieved information is then used to make an inference about the occurrence of the test word in the study list (i.e., "I remember solving an anagram for this word and I solved anagrams at study, and so this word must have been in the study list").

A postretrieval symbolic inference process is certainly compatible with introspection and the reports from participants. In addition, there are a substantial number of experimental findings that can parsimoniously be explained if, on some occasions, participants retrieve symbolic information that is then used to make an inference about whether an item was present in an experiment or list (for reviews, see Clark & Gronlund, 1996; Humphreys & Bain, 1983; Mandler, 1980). However, the subsymbolic inferential process in BCDMEM is a powerful computational tool. It produces a finely graded source of information, whereas the retrieval of symbolic information will be less finely graded. It also allows weak information to be used in a sensible way, whereas the failure to retrieve symbolic information may provide little direction to action. Most important, it allows participants to use information about what they expect to remember (e.g., a word that was read or a word that was heard) early rather than late in the inferential process.

A subsymbolic inferential process may also be compatible with much of the introspective evidence. That is, the lack of meaning in the retrieved context vector, with meaning being supplied only by the match with the reinstated context vector, seems compatible with an undifferentiated feeling of familiarity. In this case, however, it is a feeling of familiarity that is specific to the memory that the participant is expecting. That is, if participants are attempting to identify words that were read, then read words will tend to be more familiar than heard words. However, if they are attempting to identify words that they heard, then heard words will be more familiar than read words. In addition, it is a simple step to assume that at times the process that is used to retrieve the context vector continues until symbolic information becomes available. Such information could increase confidence in the correctness of a response, and in some circumstances (e.g., when it can be used to reject a distractor) it would increase sensitivity.

Approximating a Likelihood Calculation

The ubiquitous nature of the mirror effect suggests that the memory system is using both evidence that a word appeared in the study episode and evidence that it did not appear to calculate something akin to a likelihood ratio. Furthermore, several current models, including BCDMEM, propose rules that are optimal in a Bayesian sense, given their architectural constraints, and it seems that optimality will provide a useful soft constraint on future models.

Context Noise Versus Item Noise

The item noise approach provides compelling explanations for the list length effect and for word similarity effects on false recognitions. In our opinion, its main weakness is closely linked to this strength. That is, by assuming that both recall and recognition are item noise processes, it is compelled to treat them as being highly similar. There have always been some problems with this assumption. For example, word frequency has different effects in recall and recognition, and this has posed a challenge to the item noise models (see Gillund & Shiffrin, 1984). It is also known that recall and recognition respond very differently to cumulative proactive interference (Postman & Keppel, 1977).

However, the first major challenge for this approach came with Ratcliff et al.'s (1990) finding of a null list strength effect. This finding caused severe problems for the global matching models that used composite memories. Furthermore, although SAM could be altered to accommodate this finding, the result was certainly an increase in complexity. In fact, it was not until new models (REM and M&C) were introduced that a relatively parsimonious explanation for the null list strength effect was available. We have now conducted two studies in which we could have found a list length effect. None of the comparisons approached significance, and all of the effects were quite small. This finding is a significant problem for the item noise approach because to model such small effects it is necessary to assume very low levels of item similarity, making it difficult to simultaneously fit list length effects in recall and recognition.

With respect to the effect of word similarity on false recognitions, the problem faced by the item noise models is the large variability found in false recognitions. Until this variation can be explained, it will be difficult to determine whether false recognitions provide substantial support for the item noise account or whether they can be more adequately explained by the variety of mechanisms we have discussed.

Conclusion

The primary purpose of this article has been to create a modern version of a context noise model that can provide a counterpoint to the well-developed item noise and dual-processing approaches. BCDMEM provides an adequate fit to most of the standard data on which the item noise models have been validated. More important, this exercise has resulted in the identification of areas that require additional research effort. The existing literature does not give us an adequate picture of the magnitude of the list length effect or the increment in the false alarm rate produced by the study of a "similar" word. In addition, we need studies, especially modeling studies, comparing list length effects in recognition and recall. We also need to closely examine the possibility that false alarm rates for similar items are produced when the lure is generated at study or when the criterion changes on a categorical basis.

As we indicated earlier, REM has both item noise and context noise features, although to date the item noise features have been emphasized. Although coming from a different tradition, the models proposed by J. R. Anderson, Bothell, Lebiere, and Mantessa (1998) and Reder et al. (in press) also have both item noise and context noise features. Similarly, it would be possible to incorporate some item noise into BCDMEM by using a distributed instead of a local representation in the binding layer. However, we chose to present a pure context noise model to focus attention on whether recognition is primarily an *item noise* or *context noise process*. Although additional research on list length and similarity effects will be required to resolve this issue, we believe that context noise must be considered a viable alternative to the dominant item noise account.

In developing BCDMEM, we borrowed aspects from the dual-processing and source monitoring approaches. All three approaches share the use of a retrieved context and an emphasis on the importance of how the words in an experiment are processed. In addition, BCDMEM and the dual-processing approach share the use of discounting. Moreover, we made heavy use of the process dissociation procedure developed within the dual-processing approach. However, BCDMEM's success in addressing a wide range of data from both the item noise and dual-processing domains highlights the limitations of the dual-processing approach as it is currently instantiated. Unless the dual-processing approach incorporates at least some rudimentary ideas about bindings, cues, and decisions, there will be large amounts of data to which it cannot be applied. For example, a finding that there was a list length effect in recall but not in recognition would be extremely important because it would help to explain the differential susceptibility to proactive interference and the intuitive feeling that recognition was simpler than recall. However, such a finding now appears to be beyond the purview of the dual-processing approach.

References

Anderson, J. A., & Bower, G. H. (1972): Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Anderson, J. A., & Bower, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition, 2,* 406–412.

Anderson, J. R., Bothell, D., Lebiere, C., & Mantessa, M. (1998). An integrated theory of recognition memory. *Journal of Memory and Language, 38,* 341–380.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96,* 703–719.

Anisfeld, M., & Knapp, M. E. (1968). Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology, 77,* 171–179.

Bowles, N. L., & Glanzer, M. (1983). An analysis of interference in recognition memory. *Memory & Cognition, 11,* 307–315.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review, 106,* 160–179.

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Towards unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 123,* 137–160.

Chalmers, K. A., & Humphreys, M. S. (1998). Role of generalized and episode specific memories in the word frequency effect in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 610–632.

Clark, S. E. (1999). Recalling to recognize and recognizing recall. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin model* (pp. 215–244). Mahwah, NJ: Erlbaum.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33A,* 457–505.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Dennis, S. (1995). *The Sydney Morning Herald word database* [On-line]. Available: http://psy.uq.edu.au/CogPsych/Noetica

Dennis, S., & Humphreys, M. S. (1998). Cuing for context: An alternative to global matching models of recognition memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition cuing for context: An alternative to global matching models of recognition memory* (pp. 109–127). Oxford, England: Oxford University Press.

Dewhurst, S. A., & Anderson, A. J. (1999). Effects of exact and category repetition in true and false recognition memory. *Memory & Cognition, 27,* 665–673.

Dodson, C. S., & Johnson, M. K. (1996). Some problems with the process dissociation approach to memory. *Journal of Experimental Psychology: General, 125,* 181–194.

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

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

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16.

Glanzer, M., Adams, J. K., Iverson, G., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546–567.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 21–31.

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

Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1335–1369.

Grossman, L., & Eagle, M. (1970). Synonymity, antonymity, and association in false recognition responses. *Journal of Experimental Psychology, 83,* 244–248.

Gruppuso, V., Lindsay, D. S., & Kelly, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 259–278.

Hall, M. (1996). *Process dissociation procedure: The effects of manipulating interlist discriminability on estimates of familiarity.* Unpublished thesis, School of Psychology, University of Queensland, Brisbane, Queensland, Australia.

Hintzman, D. L. (1984). Minerva 2: A simulation model of human memory. *Behavior Research Methods, Instruments, and Computers, 16,* 96–101.

Hintzman, D. L., Curran, T., & Oppy, B. (1992). Effects of similarity and repetition on memory: Registration without learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 667–680.

Howard, M. W., & Kahana, M. J. (1998). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1–19.

Howard, M. W., & Kahana, M. J. (1999, August). *A distributed representation of temporal context.* Paper presented at the 32nd Annual Meeting of the Society for Mathematical Psychology, Vanderbilt University, Nashville, TN.

Humphreys, M. S., & Bain, J. D. (1983). Recognition memory: A cue and information analysis. *Memory & Cognition, 11,* 583–600.

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

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix and TODAM models. *Journal of Mathematical Psychology, 33,* 36–67.

Humphreys, M. S., Wiles, J., & Dennis, S. (1994). Towards a theory of human memory: Data structures and access processes. *Behavioral and Brain Sciences, 17,* 655–692.

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

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

MacLeod, C. M., & Nelson, T. O. (1976). A nonmonotonic lag function for false alarms to associates. *American Journal of Psychology, 89,* 127–135.

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect effects of stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 539–559.

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

Mandler, G., Pearlstone, Z., & Koopmans, H. J. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 8,* 410–423.

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 success and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review, 106,* 398–405.

Mulligan, N. W., & Hirshman, E. (1997). Measuring the bases of recognition memory: An investigation of the process dissociation framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 280–304.

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

Murdock, B. B., & Kahana, M. J. (1993a). Analysis of the list strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 689–697.

Murdock, B. B., & Kahana, M. J. (1993b). List-strength and list-length effects: Reply to Shiffrin, Ratcliff, Murnane, and Nobel. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1450–1453.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 855–874.

Ohrt, D. D., & Gronlund, S. D. (1999). List-length effect and continuous memory: Confounds and solutions. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin model* (pp. 105–126). Mahwah, NJ: Erlbaum.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review, 91,* 281–293.

Postman, L. (1951). The generalization gradient in recognition memory. *Journal of Experimental Psychology, 42,* 231–235.

Postman, L., & Keppel, G. (1977). Conditions of cumulative proactive interference. *Journal of Experimental Psychology: General, 106,* 376–403.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). The list-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single process theories, and recognition memory. *Journal of Experimental Psychology: General, 124,* 352–374.

Reder, L. M., Nhouyvansivong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (in press). 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.*

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

Schulman, A. I. (1974). The declining course of recognition memory. *Memory & Cognition, 2,* 14–18.

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 267–287.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). The list-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195.

Shiffrin, R. M., Ratcliff, R., Murnane, K., & Nobel, P. (1993). TODAM and the list-strength and list-length effects: Comment on Murdock and Kahana (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1445–1449.

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

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11,* 1–74.

Underwood, B. J. (1978). Recognition memory as a function of length of study list. *Bulletin of the Psychonomic Society, 12,* 89–91.

Underwood, B. J., & Humphreys, M. S. (1979). Context change and the role of meaning in word recognition. *American Journal of Psychology, 92,* 577–609.

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.

Appendix A

Adding the Contextual Reinstatement Parameter to the Decision Rule

In this appendix, we outline how the contextual reinstatement parameter can be added to the expression for the likelihood ratio. The contextual reinstatement parameter is the probability that an element of the context that was a one at study is a zero in the reinstated context.

Let c be the study context and c' the reconstructed study context. Let m be the retrieved memory vector, p the context noise parameter, r the learning parameter, s the sparsity, d the contextual reinstatement parameter, and n_{jk} the number of components in which $c_i' = j$ and $m_i = k$.

$$\begin{aligned}
&P(c_i' = 0 \ \& \ m_i = 0|\text{old}) \\
&= P[(c_i' = 0 \ \& \ m_i = 0 \ \& \ c_i = 0) \wedge (c_i' = 0 \ \& \ m_i = 0 \ \& \ c_i = 1)|\text{old}] \\
&= P(c_i' = 0 \ \& \ m_i = 0 \ \& \ c_i = 0|\text{old}) + P(c_i' = 0 \ \& \ m_i = 0 \ \& \ c_i = 1|\text{old}) \\
&= P(c_i' = 0 \ \& \ m_i = 0|c_i = 0 \ \& \ \text{old})P(c_i = 0|\text{old}) \\
&\quad + P(c_i' = 0 \ \& \ m_i = 0|c_i = 1 \ \& \ \text{old})P(c_i = 1|\text{old}) \\
&= (1 - p)(1 - s) + d(1 - r)(1 - p)s. \qquad \text{(A1)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' = 0 \ \& \ m_i = 0|\text{new}) \\
&= P(c_i' = 0 \ \& \ m_i = 0|c_i = 0 \ \& \ \text{new})P(c_i = 0|\text{new}) \\
&\quad + P(c_i' = 0 \ \& \ m_i = 0|c_i = 1 \ \& \ \text{new})P(c_i = 1|\text{new}) \\
&= (1 - p)(1 - s) + d(1 - p)s. \qquad \text{(A2)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' = 0 \ \& \ m_i = 0|\text{old})/P(c_i' = 0 \ \& \ m_i = 0|\text{new}) \\
&= [1 - s + d(1 - r)s]/(1 - s + ds). \qquad \text{(A3)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' = 1 \ \& \ m_i = 0|\text{old}) \\
&= P(c_i' = 1 \ \& \ m_i = 0|c_i = 0 \ \& \ \text{old})P(c_i = 0|\text{old}) \\
&\quad + P(c_i' = 1 \ \& \ m_i = 0|c_i = 1 \ \& \ \text{old})P(c_i = 1|\text{old}) \\
&= (1 - d)(1 - r)(1 - p)s. \qquad \text{(A4)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' = 1 \ \& \ m_i = 0|\text{new}) \\
&= P(c_i' = 1 \ \& \ m_i = 0|c_i = 0 \ \& \ \text{new})P(c_i = 0|\text{new}) \\
&\quad + P(c_i' = 1 \ \& \ m_i = 0|c_i = 1 \ \& \ \text{new})P(c_i = 1|\text{new}) \\
&= (1 - d)(1 - p)s. \qquad \text{(A5)}
\end{aligned}$$

$$P(c_i' = 1 \ \& \ m_i = 0|\text{old})/P(c_i' = 1 \ \& \ m_i = 0|\text{new}) = (1 - r). \qquad \text{(A6)}$$

$$\begin{aligned}
&P(c_i' = 0 \ \& \ m_i = 1|\text{old}) \\
&= P(c_i' = 0 \ \& \ m_i = 1|c_i = 0 \ \& \ \text{old})P(c_i = 0|\text{old}) \\
&\quad + P(c_i' = 0 \ \& \ m_i = 1|c_i = 1 \ \& \ \text{old})P(c_i = 1|\text{old}) \\
&= p(1 - s) + d(r + p - rp)s. \qquad \text{(A7)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' = 0 \ \& \ m_i = 1|\text{new}) \\
&= P(c_i' = 0 \ \& \ m_i = 1|c_i = 0 \ \& \ \text{new})P(c_i = 0|\text{new}) \\
&\quad + P(c_i' = 0 \ \& \ m_i = 1|c_i = 1 \ \& \ \text{new})P(c_i = 1|\text{new}) \\
&= p(1 - s) + dps. \qquad \text{(A8)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' = 0 \ \& \ m_i = 1|\text{old})/P(c_i' = 0 \ \& \ m_i = 1|\text{new}) \\
&= [p(1 - s) + d(r + p - rp)s]/[p(1 - s) + dps]. \qquad \text{(A9)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' = 1 \ \& \ m_i = 1|\text{old}) \\
&= P(c_i' = 1 \ \& \ m_i = 1|c_i = 0 \ \& \ \text{old})P(c_i = 0|\text{old}) \\
&\quad + P(c_i' = 1 \ \& \ m_i = 1|c_i = 1 \ \& \ \text{old})P(c_i = 1|\text{old}) \\
&= (1 - d)(r + p - rp)s. \qquad \text{(A10)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' = 1 \ \& \ m_i \\
&= 1|\text{new}) = P(c_i' = 1 \ \& \ m_i = 1|c_i = 0 \ \& \ \text{new})P(c_i = 0|\text{new}) \\
&\quad + P(c_i' = 1 \ \& \ m_i = 1|c_i = 1 \ \& \ \text{new})P(c_i = 1|\text{new}) \\
&= (1 - d)ps. \qquad \text{(A11)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' = 1 \ \& \ m_i = 1|\text{old})/P(c_i' = 1 \ \& \ m_i = 1|\text{new}) \\
&= (r + p - rp)/p. \qquad \text{(A12)}
\end{aligned}$$

$$\begin{aligned}
&P(c_i' \ \& \ m_i|\text{old})/P(c_i' \ \& \ m_i|\text{new}) \\
&= \{[1 - s + d(1 - r)s]/(1 - s + ds)\}^{n00}(1 - r)^{n00} \\
&\quad \{[p(1 - s) + d(r + p - rp)s]/[p(1 - s) + dps]\}^{n01} \\
&\quad [(r + p - rp)/p]^{n11}. \qquad \text{(A13)}
\end{aligned}$$

(Appendixes continue)

Appendix B

List Length Experiment 1

Method and Design

Participants

One hundred two undergraduate students from the University of Queensland participated in the experiment for course credit. A 2 × 2 design was used. The variables were length of list (short [24 words] or long [72 words]) and study–test lag (start of study session [990 s] or end of study session [480 s]). Length of list was a between-subjects variable. Study–test lag was a within-subject variable for the long lists and a between-subjects variable for the short lists. Thirty participants were allocated to each of the short conditions, and 42 were allocated to the long condition.

Stimuli

The study words were five-letter words chosen from the *Sydney Morning Herald* word database (Dennis, 1995) and fell in the range of 20–50 occurrences per million. Items were assigned to conditions randomly on a per-participant basis.

Procedure

Each condition involved a series of puzzle and rating tasks followed by an unanticipated recognition test. Both the puzzle task and the rating task were explained to participants before they began the first list so that no breaks would be necessary between tasks. The puzzle involved rearranging tiles to form a geometric pattern. The rating task involved scoring the pleasantness of a word on a 6-point scale. At the start of each rating block, participants were given a 3-s warning period to orient to the rating task. Each word was presented for 3 s, and participants were instructed that should they miss a word, they should just continue with the next word. Each rating block required the scoring of 24 words.

Table B1 shows the order and durations of the tasks. As outlined in the main text, puzzle activity was interspersed in the short list conditions to equate study–test lag for either the early words in the long list or the late words. Furthermore, all study blocks were of equal size and were separated by at least 3 min of puzzle activity.

All test blocks contained 20 old words and 20 new words. In the short lists, the old words were the final 20 words of the study list. The first 4 words were discarded to avoid any difficulties participants may have experienced in orienting to the rating task. In the long lists, 10 old words were drawn from the final 20 positions of the first study block and 10 were drawn from the final 20 positions of the last study block. The design was constructed in this way so that participants would be forced to recognize items from the entire study session (i.e., all three lists) rather than isolating the specific study block from which their words were drawn. The test was self-paced.

Results

See the main text for a description of the results.

Table B1
Timing of Component Tasks

Long list	Short list at start	Short list at end
Puzzle (180 s)	Puzzle (180 s)	
Rate 24 words (75 s)	Rate 24 words (75 s)	
Puzzle (180 s)		Puzzle (690 s)
Rate 24 words (75 s)		
Puzzle (180 s)	Puzzle (990 s)	
Rate 24 words (75 s)		Rate 24 words (75 s)
Puzzle (480 s)		Puzzle (480 s)
Test Blocks 1 and 3	Test Block 1	Test Block 1

Appendix C

List Length Experiment 2

Method and Design

Participants

Forty-eight undergraduate students from the University of Queensland participated in the experiment for course credit. A 3 × 2 design was used. The variables were list type (short [40 words], mixed [40 unique words and 80 presentations], or long [80 words]) and word frequency (high or low). List type was a within-subject variable (counterbalanced for order). Word frequency was a between-subjects variable.

Stimuli

The study words were five-letter words chosen from the *Sydney Morning Herald* word database (Dennis, 1995). High-frequency words fell in the range of 100–200 occurrences per million, whereas low-frequency words were below 4 occurrences per million. Items were assigned to conditions randomly on a per-participant basis.

Procedure

Each condition involved a series of puzzle and rating tasks followed by an intentional recognition test. Both the puzzle task and the rating task were explained to participants before they began the first list so that no breaks would be necessary between tasks. The puzzle involved rearranging tiles to form a geometric pattern. The rating task involved scoring the pleasantness of a word on a 6-point scale. Each word was presented for 3 s, and participants were instructed that should they miss a word, they should just continue with the next word.

In the short condition (AB), participants were presented with two blocks of 20 words without a break. In the long condition (ABCD), participants were presented with four blocks of 20 words. In the mixed condition

(ABBB), participants were presented with four blocks of 20 words in which the second block was repeated three times. In all cases, the retroactive design was used. In the short condition, 360 s of puzzle activity followed the last study word; in both the long condition and the list strength condition, 240 s of puzzle activity followed the presentation of the last word to equate for lag. Half of the participants received lists of high-frequency words, and half received lists of low-frequency words. The test list consisted of 20 Block 1 words, 20 Block 2 words, and 40 new words and was self-paced.

Results

See the main text for a description of the results.

Appendix D

Strength and the Process Dissociation Procedure

Method and Design

Participants

Thirty undergraduate students from the University of Queensland participated in the experiment for course credit. A 3 × 3 × 2 factorial design was used. The variables were number of presentations in List 1 (0, 1, or 4), number of presentations in List 2 (0, 1, or 4), and the list that was cued (List 1 or List 2). All variables were within-subject variables.

Stimuli

The study words were five-letter words chosen from the *Sydney Morning Herald* word database (Dennis, 1995) and fell in the range of 5–10 occurrences per million. Six words were assigned randomly in each of the study presentations. Three of these words were tested under instructions cuing for List 1, whereas the other three were tested with List 2 instructions. A second block was constructed with the same methodology to increase the number of observations per cell.

Procedure

Each of the two blocks consisted of the presentation of two study lists followed by two test lists. In the study lists, participants were asked to rate each word for pleasantness on a 5-point scale in which 1 indicated that the word was very unpleasant, 3 indicated neutrality, and 5 indicated that the word was very pleasant. Each study list contained 44 unique words and 110 presentations, and participants had 3 s to make their decision. A 5-min distractor activity that involved solving a sliding geometric puzzle was interposed between each of the study and test lists.

One of the tests involved cuing for List 1, and the other involved cuing for List 2. The order in which the tests were conducted was counterbalanced. Participants were told that the words they were about to see occurred

Table D1
Numbers of Words and Presentations Used

List 1			List 2			
Number of words	Repetitions	Presentations	Number of words	Repetitions	Presentations	Presentations in each test
0	0	0	0	0	0	11
22	1	22	0	0	0	11
22	4	88	0	0	0	11
0	0	0	22	1	22	11
0	0	0	22	4	88	11

Table D2
Mean Probability of Yes Response as a Function of Presentation List, Number of Presentations, and Instructions

	Inclusion				Exclusion			
	1 presentation		4 presentations		1 presentation		4 presentations	
Presentation list	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
List 1	.677	.020	.807	.019	.565	.025	.631	.029
List 2	.708	.023	.825	.031	.446	.028	.347	.036

Note. The false alarm rate when List 1 was cued was 0.207 (SE = 0.041), and the rate for List 2 was 0.174 (SE = 0.041). There was no significant difference between these rates, $F(1, 47) = 2.07$, $MSE = 0.0127$, $p = .157$. Recall that the experiment involved the Yonelinas (1994) procedure, so the inclusion rate was the probability that participants responded "old" to words from the target list. The exclusion rate was the probability that participants responded "old" to words from the nontarget list.

(*Appendixes continue*)

either not at all or on one list and not the other. They were to respond yes only if they saw the word on the cued list. The recognition tests were self-paced. There were 55 words on each test list. Table D1 outlines the different types of words that were tested and the numbers of presentations.

Results

Table D2 shows the means for all cells. See the main text for a description of the results.

Appendix E

Diffuseness of Retrieved Context

Method

Participants

All participants were undergraduate students from the University of Queensland who took part in the experiment for course credit. The first experiment involved 23 participants in each of the two conditions (inclusion and exclusion), and the second experiment involved 25 participants in each of the same two conditions.

Stimuli

The study words were five-letter words chosen from the *Sydney Morning Herald* word database (Dennis, 1995) and fell in the range of 100–800 occurrences per million.

Procedure and Design

The design involved three study lists presented in immediate succession, followed by a recognition test. Different rating tasks were used for each list, and these tasks were self-paced. Of the 32 List 1 words, 8 were presented once, 8 were presented twice, and 8 were presented once in List 1 and then repeated in List 2. Of the 32 List 2 words, 8 were presented once, 8 were presented twice, and 8 were repeated from List 1. All 40 List 3 words were presented once in that list only. The test list contained 88 words in total: 8 new words, 8 words presented once in List 1, 8 words presented twice in List 1, 8 words presented once in List 2, 8 words presented twice in List 2, 8 words presented once in List 1 and once in List 2, and 40 words presented once in List 3. The test words were presented one at a time, and the participant had to make a yes–no response using the computer mouse before the next test word was presented.

Results

See the main text for a description of the results.

Received November 4, 1997
Revision received August 18, 2000
Accepted August 18, 2000 ■