

The detection model of recognition using know and remember judgments

CHICAKO INOUE and FRANCIS S. BELLEZZA
Ohio University, Athens, Ohio

The signal detection model for *know* and *remember* recognition judgments was tested in two experiments. In Experiment 1, two predictions of the model were tested: (1) that measures of memory sensitivity, A' , are equivalent in value when based on either the recognition (know or remember) criterion or on the remember criterion; and (2) that there is a positive correlation between recognition bias and the proportion of know judgments that are hits, but no correlation between recognition bias and proportion of remember hits (Donaldson, 1996). Both predictions were supported by the data. In Experiment 2, the context of test items was manipulated to make it more or less similar to learning context. The detection model requires that memory sensitivity be the same for both recognition and remember judgments, regardless of test context. Alternatively, if remember judgments reflect only the retrieval of episodic information from memory, the two measures of memory sensitivity should become more disparate in value as learning and test context are made more similar. Memory sensitivity was generally the same in value for recognition and remember criteria but different across context conditions, thus supporting the detection model. The nature of the memory continuum used in detection theory is also discussed.

There seems to be a good deal of empirical support for the notion that *know* and *remember* recognition judgments represent different conscious states of recognition memory (see Gardiner, 1988; Tulving, 1985). Some experimental variables have been found to increase remember but not know judgments (Gardiner, 1988; Gardiner & Java, 1990; Gardiner & Parkin, 1990; Java, Kaminska, & Gardiner, 1995; Jones & Roediger, 1995). Still others increase know but not remember judgments (LeCompte, 1995; Rajaram, 1993). In addition to this evidence for single dissociation, there is also evidence for double dissociation. There are variables that increase remember but decrease know judgments (Conway & Dewhurst, 1995; Dewhurst & Conway, 1994; Gardiner & Java, 1990; Jones & Roediger, 1995; Parkin & Walter 1992; Rajaram, 1993; Yonelinas & Jacoby, 1995).

Not all models of memory utilize two conscious states to explain recognition performance, however (Ratcliff, Van Zandt, & McKoon, 1995). Some models require only one memory system (Gillund & Shiffrin, 1984; Glanzer, Adams, Iverson, & Kim, 1993; Hintzman, 1986; Murdock, 1982). Such a model is the signal detection model (Banks, 1970; Egan, 1958; Lockhart & Murdock, 1970). This model proposes that two different cognitive components,

a memory structure and a decision process, mediate recognition. The rationale for explicitly including a decision mechanism is that no memory model can be formulated without one (Murdock, 1974). The detection model has been proposed to account for remember and know judgments (Bellezza & Inoue, 1993; Donaldson, 1996; Hirschman & Master, 1997).

Using Signal Detection Theory With Remember and Know Judgments

Use of the detection model to explain remember and know recognition judgments is diagrammed in Figure 1. Each item is represented on a single continuum that can be thought of as the amount of item information stored in memory. The distribution of *new* test items is displayed on the left in Figure 1. Their mean strength is lower than that of the *old* items, whose distribution is displayed on the right. For purposes of discussion, we can assume the distributions to be normal with equal variances, though it is common in recognition studies for the old items to be somewhat more variable in strength than the new items (Ratcliff, Sheu, & Gronlund, 1992). The distances between the means is denoted as d' (Swets, 1986). To account for recognition judgments, we assume that the person tested places Criterion A on this memory dimension. Any test item with a value above Criterion A is responded to as an old item and is recognized. An item whose memory value is below Criterion A is responded to as a new item. The area of the distribution of old items above Criterion A indicates the hit rate, and the area of the distribution of new items above Criterion A indicates the false alarm rate. The criterion value represents a person's response bias—that is, the bias to respond as recognizing an item. If the criterion is high in value, bias is conservative. If low, bias is liberal.

Parts of this article are based on the doctoral comprehensive examinations of the first author working under the supervision of the second author. Some of its content was presented at the meeting of the Psychonomic Society, Washington, DC, November 1993. Correspondence concerning this article should be addressed to F. S. Bellezza, Department of Psychology, Ohio University, Athens, OH 45701 (e-mail: bellezza@oak.cats.ohiou.edu).

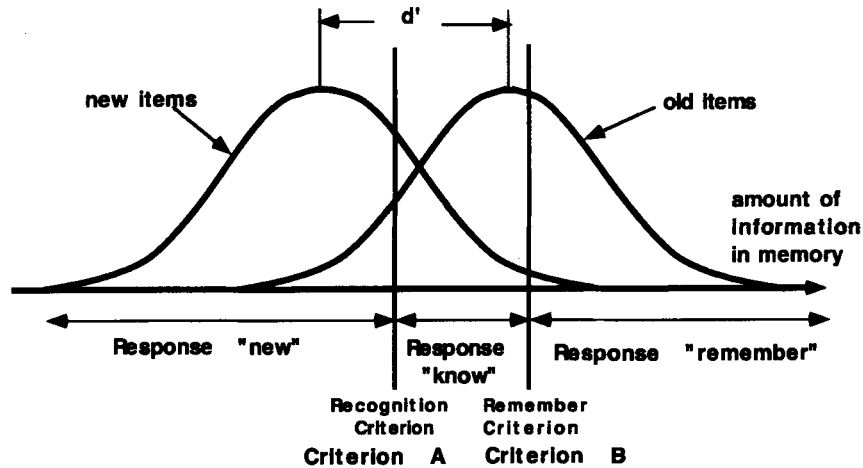


Figure 1. The remember and know criteria for the distributions of old and new items. Normal distributions are assumed; d' is a measure of memory sensitivity.

For the detection model to account for remember and know judgments, a second criterion, Criterion B, is placed above Criterion A on the memory dimension (Bellezza & Inoue, 1993; Donaldson, 1996; Hirshman & Master, 1997). Not only is any item whose strength is above Criterion B recognized, but the event of studying the item is judged to be recollected. Hence, for these items, a remember judgment is given in addition to the recognition judgment. For items whose strength is between Criteria A and B, a know judgment is made. Some recognition foils will have large amounts of information available in memory by chance, and these will elicit false positive remember or false positive know judgments. For the reasons outlined above, we refer to Criterion A as the *recognition criterion* and to Criterion B as the *remember criterion*. Hence, remember and know judgments can be dealt with in detection theory by assuming that they are mediated by two judgment criteria, much as multiple criteria define confidence ratings in signal detection theory (Green & Swets, 1966, pp. 40–43).

The power of the remember-know recognition paradigm is that the patterns of hits and false alarms are typically different for the remember and know judgments, and, as indicated above, experimental variables have different effects on these two judgments. This seems to provide compelling evidence for two independent memory systems. Yet, a complete experimental dissociation can be demonstrated by using the detection model, which is based on a single memory dimension. Depending on the placement of the criteria, the number of remember judgments can be small and the number of know judgments large, or vice versa. Also, if the criteria move on the continuum, the patterns of hits and false alarms change in different ways for remember and know judgments. Examples of precisely how this can occur are given by Donaldson (1996) and by Hirshman and Master (1997).

Testing the Detection Model

To assess the worth of the detection model, Donaldson (1996) computed memory sensitivity and bias measures

for previously published data from 28 different experiments involving know and remember judgments. He found support for the detection model. For example, smaller retention intervals resulted in larger memory sensitivity values, suggesting that the rememberer can more easily recognize old items when the retention interval is short (Gardiner & Java, 1991; Tulving, 1985). A number of other experimental manipulations also had the expected effects: Greater memory sensitivity values occurred following a word generation task as opposed to a word-reading task (Gardiner, 1988). Larger sensitivity values were found for the low-frequency word conditions than for the high-frequency word conditions (Gardiner & Java, 1990), as described by the word frequency effect. The divided attention condition of Gardiner and Parkin (1990) resulted in smaller d' values than did the undivided attention condition. Dividing attention made the subsequent recognition of items more difficult. These results suggest that the parameter values of the detection model are influenced by experimental manipulations in a predictable manner, and this lends credibility to the notion that the detection model may be useful for analyzing know and remember recognition data.

Although the values of memory sensitivity may make sense when fitted to know and remember data, why should the detection model be of value? The main reason is that it can describe remember and know data in a relatively simple manner. Only one memory system is utilized rather than two. Furthermore, there are tests of the detection model, based on two specific predictions. First, if the know and remember judgments are derived from two criteria along a single dimension, then a signal detection analysis using either criterion should produce the same value of d' , the distance between the means of the old and new items. Consider a hypothetical Case 1. A subject responds to target items with a remember judgment .40 of the time; a know judgment, .30 of the time; and a new judgment, .30. For new items, the proportions are .10, .20, and .70, respectively. If we use Criterion B in Figure 1, the hit and false alarm rates are .40 and .10, respectively, for remem-

ber judgments. This results in a memory sensitivity value of $d' = 1.03$. If we use the criterion for recognition judgments, Criterion A, the hit and false alarm rates are (.40 + .30) and (.10 + .20), respectively, resulting in the parameter value of $d' = 1.05$. These d' values are reasonably close in magnitude, as the detection model predicts.

It should be noted, however, that the model does not necessarily fit any know and remember data. Consider a hypothetical Case 2. The proportions of remember, know, and new judgments for target items are .30, .50, and .20, respectively, and the respective proportions for new items are .20, .10, and .70. The value of d' based on the remember criterion is .32, and the value based on the recognition criterion is 1.37. This kind of difference in values would not support the detection model.

Using published data from 28 experiments in order to test the prediction of equality of d' 's, Donaldson (1996) found a mean value of 1.80 for d' based on remember criteria and a value of 1.71 for d' based on recognition criteria—that is, based on the proportions of remember and know judgments combined. Using a nonparametric measure of memory sensitivity, A' , which does not assume normal distributions of memory strengths with equal variance for old and new items, Donaldson (1996) found mean A' values of .83 for the remember criteria and .86 for the recognition criteria. He concluded that these mean memory sensitivity values are close enough in value to support the detection model.

One might suppose that a second prediction for the model is that the recognition criterion will always be lower in value than the remember criterion. However, because the recognition criterion is based on calculations combining the proportions of know and remember judgments, it always has more of the distributions of new items and old items above it than does the remember criterion, which is based only on the remember judgments. Hence, the value of any bias measure, such as C (Macmillan & Creelman, 1990), will always be lower for the recognition criterion than for the remember criterion, except when there are only know or only remember judgments.

There exists, however, a second prediction that can be made from the detection model. Donaldson (1996) has pointed out that if the recognition and remember criteria are relatively high in value, as shown in Figure 1, there will be more know judgments that are hits than there are know judgments that are false alarms. This is because the region between the two criteria will contain a large area of the distribution of old items and a smaller area of new items. Alternatively, if the two criteria are low in value—that is, positioned on the left side of the continuum—the region between them will contain a large area of the distribution of new items and a smaller area of the old items. If the same reasoning is applied to the remember judgments, the results are different. The proportion of remember judgments that are hits is always larger than the proportion of remember judgments that are false alarms, regardless of the position of the criteria.

Hence, the prediction can be made that there will a positive correlation between value of the recognition criterion and the difference between hits and false alarms for the know judgments. Also, there will be no such relationship between the recognition criterion and this difference computed for remember judgments. Donaldson (1996) analyzed the data from 80 previously published comparisons of know and remember judgments and also from the data he collected from 40 individuals in a recognition experiment. In both instances, he found, as predicted, a positive correlation for the know judgments and no correlation for the remember judgments.

Hirshman and Master (1997) made a prediction from the detection model somewhat similar to that of Donaldson (1996). They proposed that the effects of experimental manipulations are always somewhat less reliable for know judgments than for remember judgments. This is because shifts in criteria placement can produce greater changes in know judgments than in remember judgments. The relative number of know judgments may vary across experimental conditions because of criteria shifts, but this will not occur for the relative number of remember judgments. Hirshman and Master reanalyzed data from seven previously published studies in which a variable was repeatedly manipulated but inconsistent in its effect on know and remember judgments. They found that in all seven studies the inconsistencies occurred only for know judgments.

In the rest of this paper, we do the following: First, we report the results of an experiment to show that remember and know judgments for individual subjects may be based on the same memory system but on different criteria. Donaldson (1996) has fitted detection parameters for individual subjects on the basis of remember and familiar judgments, but his procedure deviates from the typical remember-know procedure, as we discuss below. Second, we report data from an experiment in which context was manipulated to affect memory sensitivity. The results indicate that both remember and know judgments are mediated by episodic memory, contrary to the assumptions of dual-memory theory (Tulving, 1985), but as is expected from the detection model. Third, we discuss the nature of the memory continuum used in the detection model.

EXPERIMENT 1

The detection model seems to fit mean hit and mean false alarm rates from previously published studies reasonably well (Donaldson, 1996). However, the question can be raised whether the model fits data from individual subjects. That is, rather than compute memory sensitivity and bias from the mean hit and false alarm rates from a sample of people, the parameters can be computed for each individual, and the mean of these parameter values derived.

Donaldson (1996) reported results from such a study, but collected remember and familiar judgments instead

of remember and know judgments. Subjects first rated each test item on a 4-point rating scale of recognition before labeling their memory experience. But they reported feeling very uncomfortable about using the usual labels of *know* and *remember* after they had used the rating scale (Donaldson, 1996, p. 527). They did feel comfortable, however, with the labels *familiar* and *remember*, which were the labels then used. This procedure, however, is problematic. Because subjects were asked to provide confidence ratings before providing remember and familiar judgments, the demands of the rating task could have resulted in remember and familiar judgments being interpreted as confidence ratings also. Know and remember judgments should not be interpreted in this way. In Experiment 1, we tested students using know and remember judgments according to the standard procedure. Detection model parameters were computed for each subject, and the fit of the model was assessed.

Method

Subjects. Thirty-one Ohio University undergraduates participated for course credit.

Materials. The 20 most frequent exemplars in 10 categories (fruits, vegetables, sports, clothing, four-footed animals, body parts, vehicles, furniture, birds, and alcoholic beverages) were selected from Battig and Montague (1969). The 20 exemplars from each category were divided into two sets. One set contained 100 exemplars with even-numbered frequency rankings in the categories, and the other set contained 100 exemplars with odd-numbered frequency rankings. For 16 of the 31 subjects, the first set was used as the target words, and the second set as the foils. For the remaining subjects, the second set comprised the targets and the first set the foils.

Design and Procedure. The subjects studied the material and were tested in small groups. During the study session, the 100 items were presented in a random order, each one for 5 sec, using a transparency projector. The experimenter read each word aloud as it was shown on the screen. The students rated the pleasantness of each word on a 7-point rating scale. These ratings were not analyzed. After presentation of the target words, the students were instructed to come back 1 week later for another experiment. During the second session, however, a know versus remember recognition test was administered. Each subject was given a booklet, including an instruction sheet and three answer sheets. The written instructions, similar to those of Gardiner and Java (1990), explained the differences between the remember and know judgments. Two examples of each response were provided. Examples of the basis for conscious recollection included the physical appearance of the word, some idea that the subject thought of during its presentation, or some action that he or she performed during its presentation. If such an experience could be remembered, the subjects were told to respond that they remembered the item. If they merely knew that the word was presented, but were not aware of any of such an experience, they were asked to respond with the know judgment. Otherwise, the indicated word was unrecognized. On the test sheets the randomized list of targets and foils were printed in two columns. Next to each item, three letters (*U*, *K*, and *R*) were printed as abbreviations for *unrecognized*, *know*, and *remember*, respectively. The subjects were told to circle one of the three letters on the response sheets given them. Finally, after the recognition test, the subjects were asked to write down any two of the test words they judged as remember words and to describe the recollective experience associated with each. They were also told to write down two examples of know words.

Results

The hit rate used to determine the remember criterion was computed by taking the total number of remember hits, adding the value of 0.5 and then dividing by $N+1$ (Snodgrass & Corwin, 1988). The same procedure was used for computing the false alarm rates. The hit and false alarm rates for determining the recognition criterion were computed by combining the know and remember judgment categories. That is, any item for which a recognition judgment was given had a value above the recognition criterion.

The hit and false alarm rates, as well as the values for nonparametric measures of memory sensitivity and bias, A' and B_D'' , respectively, (Donaldson, 1992) are shown in Table 1. The reason for using the parameters A' and B_D'' rather than d' and C based on normal distributions is that the variance of the old items is typically larger in magnitude than the new. This results in d' increasing in value as the criterion increases from recognition criterion to the remember criterion (Ratcliff et al., 1992). The measure of memory sensitivity, A' , does not have this problem (Donaldson, 1993). As can be seen from Table 1, the A' values were very close in value, although their means were reliably different.

We also tested the prediction that as the recognition criterion increased in value, there would be a greater number of hits and a smaller number of false alarms occurring with the know judgments. This would not occur, however, for the remember judgments. In doing this test, we followed the procedure of Donaldson (1996), who suggested that the nonparametric signal detection parameters A' and B_D'' be used. For each subject, we correlated the value of the nonparametric recognition criterion, B_D'' , with the memory sensitivity value of A' computed from only the know judgments. The parameter A' computed for only the know judgments is a measure that is not very interpretable in terms of signal detection theory, but it reflects the relation between hits and false alarms for test items falling between the two criteria (Donaldson, 1996). It is large in value when there are many hits and few false alarms, and it is low in value when there are few hits and many false alarms. The correlation found between the recognition bias B_D'' and A' for know judgments only was .34 ($p < .05$, one-tailed test). However, the correlation between recognition bias B_D'' and A' based on remember judgments was $-.05$ and nonsignificant.

Table 1
Performance Measures From Experiment 1

Judgment	Response Proportions		Sensitivity (A')	Bias (B_D'')
	Hits	False Alarms		
Remember	.31	.14	.69	.76
Know	.38 (.69)	.27 (.41)	.72	-.16

Note—The values in parentheses represent the proportions of hits and false alarms used to determine the values of memory sensitivity and bias based on the recognition criterion.

Discussion

As expected from the detection model, the mean memory sensitivity parameters based on the recognition and remember criteria, .69 and .72, were close in value. The parameter A' varies in value from .50, representing chance responding, to 1.00, reflecting perfect recognition. Furthermore, for subjects whose recognition criterion was low in value, there was a tendency to give a relatively larger number of false alarms in their know judgments than hits, whereas those subjects whose recognition criterion was high in value generated relatively more hits than false alarms. The results of Experiment 1, therefore, lend support to the notion that know and remember judgments represent different criteria on the same memory continuum. Of special interest, however, is whether the small advantage in A' sensitivity based on the recognition criterion over the remember criterion can be replicated.

Group parameters versus individual parameters.

Detection parameters such as A' and B_D'' can be calculated from group hits and false alarm rates, as in Donaldson (1996) and below when we discuss the use of sure versus unsure judgments by Gardiner and Java (1990) and by Rajaram (1993). The question arises whether these values are similar to those obtained by averaging values of A' and B_D'' computed individually for each subject.

Macmillan and Kaplan (1985) found that the two sets of values for the parameters d' and C are generally similar for psychophysical data. We wished to determine whether detection parameters A' and B_D'' were similar for recognition data when computed from group hit and false alarm rates versus individual hit and false alarm rates. To test this, we computed parameter values individually from the hit and false alarm rates of each of the 31 students. We did this for both know and remember judgments. We then compared these group-derived parameter values with the values in Table 1, using one-sample t tests. None of the tests were significant, indicating that the values of signal detection parameters obtained were similar in value.

Relation between know and remember judgments.

Know and remember judgments have been interpreted by some investigators as representing mutually exclusive unconscious and conscious memory processes (e.g., Rajaram, 1993). However, the argument has also been made that they should be independent (Yonelinas & Jacoby, 1995) or redundant (Knowlton & Squire, 1995). In the detection model, any item that exceeds the criterion for a remember judgment automatically exceeds the criterion for a know judgment. This is analogous to the situation in signal detection theory in which any stimulus that exceeds the criterion labeled as *very confident* that the stimulus was a signal exceeds the criterion labeled *somewhat confident*. So, in the detection model, the relation between the remember and know judgments can be best described as redundancy rather than independence or exclusivity.

According to the redundancy hypothesis of Knowlton and Squire (1995), a know judgment is given if there is information in semantic memory indicating that a test item has been presented. A remember judgment is made

only if some additional amount of information is available in episodic memory. This interpretation, however, is inconsistent with the detection model. The reason is as follows: According to the Knowlton and Squire version of dual-memory theory, any manipulation of episodic information, such as increasing the similarity of learning context and testing context, should enhance the accuracy of remember judgments but not influence know judgments. Any episodic information retrieved for target items increases the target mean for remember judgments but not the target mean for know judgments. But this effect of episodic information violates the basic assumption of the detection model that memory sensitivity should be equivalent when computed with either the recognition or the remember criteria, regardless of the experimental conditions. We tested this prediction in Experiment 2.

EXPERIMENT 2

The similarity of learning and test context of items was manipulated by presenting unrelated word pairs, with the first word as the context word and the second as the target word. Any associations formed between the unrelated words were primarily the result of learning during the experiment and were, therefore, episodic rather than semantic in nature. Each target word was tested with either the same or a different context word. If the two types of recognition judgments are based on two memory systems, the context change should affect only remember judgments because contextual information of past events are episodic in nature (Knowlton & Squire, 1995; Tulving, 1985). According to the detection model, however, subjects' degree of access to item information in memory determines their recognition judgments. When the test context is different from the learning context, the amount of information available will be reduced (Feenan & Snodgrass, 1990). Thus, memory sensitivity for both know and remember judgments should be reduced.

Method

Subjects. Eighty-one undergraduate students in introductory psychology classes participated to fulfill a course requirement. The subjects were tested in groups of 20–30 individuals.

Materials and Design. Three hundred ninety words were chosen from Toggia and Battig (1978). The words were concrete nouns taken from Clusters 7 and 8, containing words with high ratings on concreteness and imagery. All words had mean imagery ratings, mean concreteness ratings, and mean familiarity ratings greater than 5.0, on the basis of a 7-point scale. A study list of 120 pairs of unrelated words was constructed. Each pair of words consisted of a context word on the left and a target word on the right (e.g., *arm–blossom*, *oats–picture*, and *vest–water*). The remaining 150 words from the original 390 words were used in the test list either as new context words or as distractor target words, as is explained below. In the recognition test, subjects gave recognition judgments only for the target words.

As is shown in Table 2, context was manipulated at the time of testing by constructing test pairs with same-context, different-context, or new-context words. For the studied target words, *same context* referred to the unchanged pairs (e.g., *arm–BLOSSOM*), *different context* referred to a target word with a context word from some other studied pair (e.g., *vest–PICTURE*), and *new context* referred to a tar-

Table 2
Examples of Study Pairs and Test Pairs

Study Pairs (Context-Target)	Context Type	Test Pairs	
		Old Target	New Foil
<i>arm-blossom</i>	same	<i>arm-BLOSSOM</i>	<i>oats-QUEEN</i>
<i>oats-picture</i>	different	<i>vest-PICTURE</i>	<i>grass-SAUCE</i>
<i>vest-water</i>			
<i>grass-flood</i>	new	<i>wine-FLOOD</i>	<i>ink-GOLD</i>

get word paired with a context word that had never been used in the studied pairs (e.g., *wine-FLOOD*). There were 30 pairs of each type to allow for testing of 90 of the original 120 target words presented.

There were also 90 test pairs that contained new target words. These were the foils (distractor items). When target words were replaced by new words, there was no functional difference between same-context foils or different-context foils, because the target words in each pair were new, though the context words were old. There were 60 foils of this type using 60 context words that had been presented but not yet been used for test pairs containing old target words, as described above (e.g., *oats-QUEEN* and *grass-SAUCE*). The remaining 30 new-context distractor pairs contained both new context words and new target words never presented before (e.g., *ink-GOLD*).

As when creating the study pairs, we tried to use unrelated words for the test pairs. Three different forms of the list of test pairs were used in order to counterbalance the context types of old targets. Old targets were assigned to the three context types an approximately equal number of times across subjects. The 180 recognition pairs were printed on test sheets in a random order. The target words appeared only on the right side of the hyphens and were capitalized to emphasize that recognition judgments should be based only on the target words, not on the pairs.

Procedure. The experiment consisted of two sessions, with a 24-h retention interval between the sessions. In each session, the subjects completed the given tasks within 45 min. During the first session, the subjects were told that they would see each of 120 unrelated word pairs presented on the screen. Each unrelated pair was presented for 15 sec by a Macintosh SE computer connected to an overhead projector. The subjects were asked to generate and write down a meaningful sentence for each pair. An audible beep signaled the presentation of each new pair. After seeing all 120 pairs, the subjects were told to return the next day and were dismissed.

On the following day, the subjects were tested, using the remember and know judgments. The procedure and instructions for the second session were similar to those used in Experiment 1 and used by Gardiner and his colleagues (e.g., Gardiner & Java, 1990). The primary difference from the previous experiments was that the recognition test was performed on target words presented with the same, different, or new context words. To ensure that the students based their recognition judgments only on the target words, they were asked to cross out the left-hand word (i.e., the context word) after looking at the pair and to make one of the three possible judgments, *U* (unrecognized), *K* (know), or *R* (remember) to the right-hand target word.

Results

The data from 5 subjects were discarded because of failure to follow instructions. The values for hits and false alarms for the recognition and remember criteria were computed individually for each subject in each of the three context conditions, using the procedure recommended by Snodgrass and Corwin (1988). Variable means for the various conditions are shown in Table 3. Shown in Fig-

ure 2 are the mean A' sensitivity values for the three context conditions.

Same context resulted in larger values of memory sensitivity for both the recognition and remember criteria than did different or new context. This result supports the assumption of detection theory that experimental manipulations should affect memory sensitivity regardless of whether the recognition judgments are labeled know or remember. The recognition versus remember criteria generally produced similar A' values in the three context conditions. However, the A' value based on the recognition criterion was significantly larger than the A' value based on the remember criterion in the different-context condition. Yet this difference was small in comparison with the effects of context manipulation.

As in Experiment 1, a correlation was computed across the 76 subjects between the value of A' computed from only the know judgments and the value of the recognition criterion, B_D'' (Donaldson, 1996). This was done separately for each of the three context conditions, using the appropriate data for each student. The correlations for the same-context, different-context, and new-context conditions were .66, .52, and .54, respectively. All correlations were significantly different from zero ($p < .001$). When the correlations were computed for the remember judgments, the respective values were .00, .24, and .12. Only the value of .24 was significantly different from zero ($p < .05$). These correlational results tend to support the detection model, which predicts a positive correlation for know judgments and no correlation for remember judgments.

Discussion

Context manipulation affected memory sensitivity values derived from both know and remember judgments. The prediction of dual-memory models that a change of context affects only remember judgments, mediated by episodic memory, and not know judgments, mediated by semantic memory (Gardiner, 1988; Knowlton & Squire,

Table 3
Summary of Performance Measures From Experiment 2

Context	Measures			
	Response Proportions		Sensitivity (A')	Bias (B_D'')
	Hits	False Alarms		
Same				
Remember	.47	.10	.79	.77
Know	.28 (.75)	.25 (.35)	.79	-.16
Different				
Remember	.28	.10	.71	.86
Know	.36 (.64)	.25 (.35)	.73	.03
New				
Remember	.26	.09	.72	.89
Know	.36 (.62)	.25 (.34)	.72	.09

Note—The values in parentheses represent the proportions of hits and false alarms used to determine the values of memory sensitivity and bias on the basis of the recognition criterion. The false alarm rates for the *same* context and the *different* context are based on the same items.

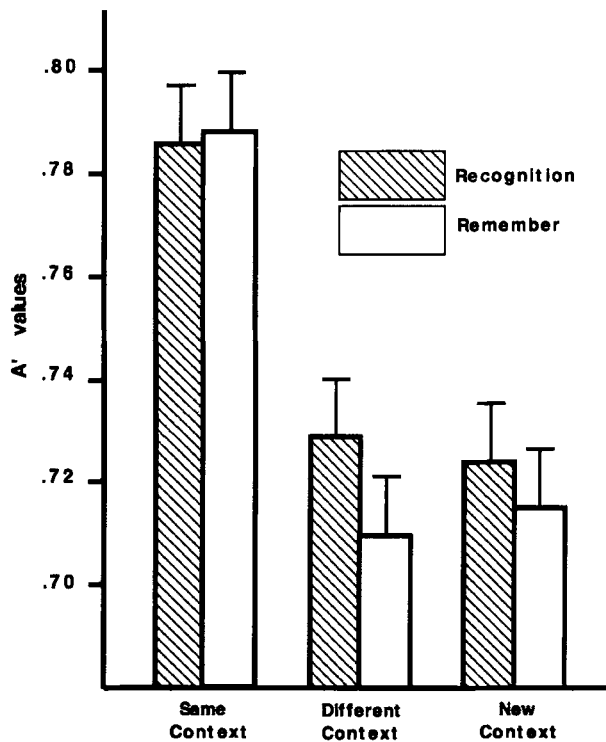


Figure 2. Mean A' values from Experiment 2 based on the recognition and remember criteria for same, different, and new context. Error bars indicate 95% confidence intervals using the within-subjects error term (Loftus & Masson, 1994).

1995; Tulving, 1985), was not confirmed. Memory sensitivity was greatest in the same-context condition because more information was made accessible by the context cues when the recognition judgment was made. This effect demonstrates the principle of encoding specificity (Tulving & Thomson, 1973).

Furthermore, memory sensitivity values for the recognition and remember criteria were similar in value within each context condition. The one significant difference found was small and not easily explained by any dual-memory theory.

As in Experiment 1, we determined whether group hit and false-alarm proportions resulted in parameter values comparable to parameter values derived from the data of individual subjects. One-sample t tests were performed on the parameter values in each of the six conditions. We compared the memory sensitivity and bias values based on the group proportions with the 76 values obtained from the individual subjects. There were no differences when the parameter A' was tested, and only one significant difference when the parameter B_D'' was tested. For remember judgments, the B_D'' value based on mean hits and false alarms in the different-context condition was .06 larger than the corresponding mean value.

These results, along with similar results from Experiment 1, suggest that detection parameters A' and B_D'' com-

puted from group data are close in value to parameter means based on individual subjects. These results lend validity to meta-analyses of detection parameters such as those reported below from Gardiner and Java (1990) and from Rajaram (1993) and for those reported by Donaldson (1996). A necessary caveat is that the correspondence between results using group versus individual hit and false alarm rates might be quite different when many of the proportions used are close to the values of 0 or 1 or when subjects vary widely in their memory sensitivity or bias (Macmillan & Kaplan, 1985).

REMEMBER AND KNOW JUDGMENTS AS CONFIDENCE JUDGMENTS

In contrast to the results of Experiments 1 and 2, other researchers have presented data which suggest that remember and know judgments cannot be interpreted in the same manner as can confidence judgments in recognition. The patterns of hits and false alarms obtained from know and remember judgments have differed from the pattern obtained by using confidence judgments such as unsure and sure. Therefore, the argument has been made that the two sets of labels are qualitatively different. For example, Gardiner and Java (1990, Experiment 2) compared words and nonwords using know and remember judgments and also utilizing the labels sure and unsure (Gardiner & Java, 1990, Experiment 3). Because the patterns of hits and false alarms were different, Gardiner and Java concluded that measures of conscious awareness are not equivalent to confidence ratings.

We performed a detection analysis, utilizing Gardiner and Java's (1990) mean hits and false alarms. The results are shown in Table 4. Inspection of the values for the detection parameters indicates that in Experiment 2 of Gardiner and Java (1990), memory-sensitivity values for nonwords and words were quite similar. When we used the unsure and sure judgments from Experiment 3, the results were somewhat different. There was greater memory sensitivity for nonwords than for words.

Comparing the bias values derived from the know and remember judgments with those from the unsure and sure judgments, we can see that the unsure and sure judgments resulted in criteria somewhat lower in value than did the know and remember judgments. This is especially true when one compares the label unsure with the label know. This difference is easily accounted for within a signal detection framework.

Rajaram (1993) presented data from recognition of items in a repetition priming condition and in a control condition in which priming occurred with a word unrelated to each target. Remember and know judgments were collected (Rajaram, 1993, Experiment 3), as well as sure and unsure judgments (Rajaram, 1993, Experiment 4). The results of our analysis of the mean hit and false alarm rates are displayed in Table 4. As in Gardiner and Java (1990), the sure and unsure judgments resulted in a

Table 4
Summary of Sensitivity Parameters Computed From the Data of
Gardiner and Java (1990) and Rajaram (1993)

	Response Proportions		Memory Sensitivity (A')	Response Bias (B_D'')
	Hits	False Alarms		
Gardiner & Java (1990, Experiment 2)				
Nonwords-remember	.19	.03	.75	.99
Nonwords-know	.30 (.49)	.12 (.15)	.77	.71
Words-remember	.28	.04	.78	.97
Words-know	.16 (.44)	.09 (.15)	.75	.76
Gardiner & Java (1990, Experiment 3)				
Nonwords-sure	.39	.07	.79	.91
Nonwords-unsure	.27 (.66)	.22 (.29)	.77	.12
Words-sure	.33	.13	.71	.86
Words-unsure	.27 (.60)	.22 (.35)	.70	.11
Rajaram (1993, Experiment 3)				
Repetition primes-remember	.43	.05	.82	.92
Repetition primes-know	.24 (.67)	.18 (.23)	.81	.24
Unrelated primes-remember	.42	.05	.82	.93
Unrelated primes-know	.18 (.60)	.13 (.18)	.80	.50
Rajaram (1993, Experiment 4)				
Repetition primes-sure	.57	.06	.86	.84
Repetition primes-unsure	.18 (.75)	.19 (.25)	.83	.00
Unrelated primes-sure	.54	.04	.86	.91
Unrelated primes-unsure	.16 (.70)	.16 (.20)	.83	.26

Note—The values in parentheses represent the proportion of hits and false alarms used to determine the values of memory sensitivity and bias based on the recognition criterion.

more liberal bias than did the respective know and remember judgments, again with the unsure criterion lower on the memory continuum than on the know criterion.

GENERAL DISCUSSION

Our data generally support a detection model explanation of remember and know judgments of recognition (Bellezza & Inoue, 1993; Donaldson, 1996; Hirshman & Master, 1997). The model is represented by a continuum of item information available in memory coupled with a flexible criterion or with multiple criteria by which the respondent decides whether the test item represents a previously experienced event. Furthermore, the respondent can report the value of the event on the continuum by using verbal labels such as know or remember. These correspond to confidence labels.

Three results support the detection model. First, memory sensitivity values for recognition and remember judgments were close in value. The statistically significant differences found in memory sensitivity for recognition versus remember judgments were nonsystematic and small in magnitude compared with the effects of context manipulation. Also, differences were small in comparison with the range of values possible, as was demonstrated in Case 2 above. Second, there was a positive correlation between the values of the recognition criterion and the values of memory sensitivity computed with know judgments only but little correlation between the recognition criterion and memory sensitivity based on remember judgments (Donaldson, 1996). Third, the con-

text manipulations accessing episodic memory in Experiment 2 affected memory sensitivity in the same manner for both know and remember criteria. Dual-memory theory suggests that sensitivity for remember judgments should have been more affected than sensitivity for know judgments.

We also have demonstrated that using mean hit and false alarm rates produced only small distortions in values for memory sensitivity and bias (Macmillan & Kaplan, 1985). This is important in the discussion of implications of studies using know and remember judgments in which only group means are available.

The Nature of the Memory Continuum

The detection model of recognition utilizes a unidimensional memory continuum that is often labeled as the strength of the test items in memory. In the traditional signal detection model (Green & Swets, 1966), the continuum is actually the likelihood that the test item is an old item rather than a new item. According to the classical signal detection model, two hypotheses are considered before a judgment is made: The null hypothesis is that the test item is a new item, and the probability that this hypothesis is true is subjectively assessed by the respondent on the basis of information in memory. The alternative hypothesis is that the test item is old, and this probability is similarly assessed by the respondent. To make a decision, a likelihood ratio is formed from the probability that the item is old over the probability that it is new. If this ratio is above the likelihood criterion, the judgment made is *old*—that is, recognized. Otherwise the

judgment made is *new*, or not recognized. It is important not to equate memory strength with neural activity or strength of the memory trace. The continuum represents the amount of information available in memory relevant to the decision being made. The likelihood ratios can be based on qualitative, quantitative, multidimensional information (Ashby, 1992), and even on information from multiple-memory systems, as long as there is a way to map the information onto one continuum. Green and Swets (1966, pp. 15–16) gave the example of the wide variety of information used by weather forecasters to predict rain versus no rain for the following day.

The likelihood procedure has been criticized as too complicated to be psychologically valid, and somewhat simpler alternatives for testing the two hypotheses have been suggested and tested (Hirshman, 1995). Nevertheless, Bayesian approaches to decision making, including the use of likelihood ratios, have been used in memory models (e.g., Anderson, 1990).

Global-activation models of memory (Gillund & Shiffrin, 1984; Hintzman, 1986; Murdock, 1982) also share many features with the detection model. Ratcliff et al. (1992) provide a description of the ways in which these models are alike. In a global activation model, the strength of each test item is determined by the degree of activation in memory produced by the test item and its context. The greater the information in memory similar to the specifics of the test situation, the greater the memory activation. The main difference between global models and the detection model is that in the activation models the recognition judgment is determined by a preset criterion that the strength of the item must exceed rather than by a likelihood ratio computed from the probabilities of the alternative hypotheses.

Factors Influencing Criteria Placement

In the detection model, recognition judgments are determined by placement of the criteria on the memory continuum. These criterion values respond to manipulations such as the probability of new versus old items (Hirshman & Henzler, 1998), task demands sometimes explicitly communicated using payoff matrices, the presence of learning context (as in Experiment 2), retention interval, and so on. Yet one of the unanswered questions regarding the model is how respondents set their criteria (Hirshman, 1995), especially when various verbal labels for these criteria are provided. From the data of Gardiner and Java (1990), it appears that the pair of verbal labels *remember* and *know* are different in value, if not in kind, from two criteria based on the labels *sure* and *unsure*. The label *unsure* represents a more liberal criterion than does the label *know*. However, it should be possible not only to measure these multiple criteria but also to manipulate them experimentally. For example, one way, though not very subtle, would be to have two different payoff matrices for the know and remember judgments.

As the payoffs change, the criterion values should also change in a predictable manner. But the memory sensitivity values should remain constant. Though we argue that different patterns of hit and false alarm rates for know and remember criteria do not, according to the detection model, demonstrate process dissociation in memory, there remains the problem of experimentally demonstrating criterion independence or, at least, demonstrating how the particular verbal labels imposed by the experimenter result in the particular bias values found for the respondents.

REFERENCES

- ANDERSON, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- ASHBY, F. G. (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.
- BANKS, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74*, 81-99.
- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, *80*(3, Pt. 2), 1-46.
- BELLEZZA, F. S., & INOUE, C. (1993, November). *Awareness as a continuum: Implications for memory research*. Paper presented at the meeting of the Psychonomic Society, Washington, DC.
- CONWAY, M., & DEWHURST, S. (1995). The self and recollective experience. *Applied Cognitive Psychology*, *9*, 1-19.
- DEWHURST, S., & CONWAY, M. (1994). Pictures, images and recollective experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 1088-1098.
- DONALDSON, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*, 275-277.
- DONALDSON, W. (1993). Accuracy of d' and A' as estimates of sensitivity. *Bulletin of the Psychonomic Society*, *31*, 271-274.
- DONALDSON, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*, 523-533.
- EGAN, J. P. (1958). *Recognition memory and the operating characteristic* (Tech. Note No. AFCRC-TN-58-51, AO-152650). Bloomington, IN: Indiana University, Hearing and Communication Laboratory.
- FEENAN, K., & SNODGRASS, J. G. (1990). The effect of context on discrimination and bias in recognition memory for pictures and words. *Memory & Cognition*, *18*, 515-527.
- GARDINER, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309-313.
- GARDINER, J. M., & JAVA, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, *18*, 23-30.
- GARDINER, J. M., & JAVA, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition*, *19*, 617-623.
- GARDINER, J. M., & PARKIN, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition*, *18*, 579-583.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.
- GLANZER, M., ADAMS, J. K., IVERSON, G. J., & KIM, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*, 546-567.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- HINTZMAN, D. (1986). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528-551.
- HIRSHMAN, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 302-313.
- HIRSHMAN, E., & HENZLER, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, *9*, 61-65.

- HIRSHMAN, E., & MASTER, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition*, **25**, 345-351.
- JAVA, R., KAMINSKA, Z., & GARDINER, J. (1995). Recognition memory and awareness for famous and obscure musical themes. *European Journal of Cognitive Psychology*, **7**, 41-53.
- JONES, T., & ROEDIGER, H. L., III (1995). The experiential basis of serial position effects. *European Journal of Cognitive Psychology*, **7**, 65-86.
- KNOWLTON, B. J., & SQUIRE, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 699-710.
- LECOMPTE, D. C. (1995). Recollective experience in the revelation effect: Separating the contributions of recollection and familiarity. *Memory & Cognition*, **23**, 324-334.
- LOCKHART, R. S., & MURDOCK, B. B., JR. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, **74**, 100-109.
- LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1**, 476-490.
- MACMILLAN, N. A., & CREELMAN, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin*, **107**, 401-413.
- MACMILLAN, N. A., & KAPLAN, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, **98**, 185-199.
- MURDOCK, B. B., JR. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.
- MURDOCK, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.
- PARKIN, A. J., & WALTER, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology & Aging*, **7**, 290-298.
- RAJARAM, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, **21**, 89-102.
- RATCLIFF, R., SHEU, C. F., & GRONLUND, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, **99**, 518-535.
- RATCLIFF, R., VAN ZANDT, T., & MCKOON, G. (1995). Process dissociation, single-process theories, and recognition memory. *Journal of Experimental Psychology: General*, **124**, 352-374.
- SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.
- SWETS, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, **99**, 100-117.
- TOGLIA, M. P., & BATTIG, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.
- TULVING, E. (1985). Memory and consciousness. *Canadian Psychology*, **26**, 1-12.
- TULVING, E., & THOMSON, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.
- YONELINAS, A. P., & JACOBY, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory & Language*, **34**, 622-643.

(Manuscript received December 17, 1996;
revision accepted for publication January 17, 1997.)