CHAPTER NINE

Recognition

A similar result comes about when a definite setting is only nascently aroused. We then feel that we have seen the object already, but when or where we cannot say, though we may seem to ourselves to be on the brink of saying it. . . . It tingles, it trembles on the verge, but does not come. Just such a tingling and trembling of unrecovered associates is the penumbra of recognition that may surround any experience and make it seem familiar, though we know not why.

—William James (1890)

People often say that although they cannot quite think of the answer, they would know it if they saw it. In other words, people believe that a recognition test is "easier" than a recall test—although, as will be seen below, this statement is often wrong. The essential difference between recall and recognition tests was described by Hollingworth (1913): In a recall test, the experimenter provides the context and the subject has to retrieve the target; in a recognition test, the experimenter provides the target and the subject has to retrieve the context.

One classic experiment illustrates why many people think recognition is "easy." Shepard (1967) presented subjects with lengthy series of stimuli, and then at test presented two stimuli. One stimulus was from the list that the subjects had just studied, and one was a similar item that had not been seen. Some subjects saw 540 words, some saw 612 sentences, some saw 1224 sentences, and some saw 612 photographs. Subjects correctly recognized 88% of the words, 89% of the 612 sentences, 88% of the 1224 sentences, and almost 100% of the pictures. Some subjects who had seen the pictures were tested one week later, and even after that length of time, they still correctly recognized 87% of the pictures. However, if the correct cues are provided for recall tests, performance can be just as impressive. For example, Mäntylä (1986) presented subjects with 504 words, and each subject generated three properties of the word. At test, the experimenter presented the properties, and the subject had to respond with the original word. When tested immediately, subjects recalled approximately 91% of the words.

Shepard's test is known as a two alternative forced choice, or 2AFC test. In many college courses, your knowledge is assessed using a 4AFC test, in which four alternatives are provided. The advantage of this technique is that the experimenter can manipulate the type and number of distractor words. This method is very useful because it can give detailed information about the type of errors people make. For example, when adults make an error, they typically select an item that is related to the correct item by meaning (Underwood & Freund, 1968), whereas third-graders typically select an item that sounds like or is acoustically related to an old item (Bach & Underwood, 1970).

A second type of recognition test is often referred to as a *yes-no recognition test*. Subjects see a series of items, and then at test are presented with a single item, which they indicate is either from the studied list or is not from the studied list. Items that were originally presented are known as old items, and items that were not shown in the study list are known as new items. The main difference between the two types of tests is what is presented at test. With the yes-no method, only one item at a time is presented at test, and subjects have to indicate whether this item is old or new (see Table 9.1). In the forced choice procedure, two (or more) items are presented, and the subject has to indicate which one item was an old item. The new items in either test are referred to as distractors or lures.

Although the yes-no method might appear simple, it raises a very complicated issue. Subjects respond by indicating whether the item is old or new, and the test item may in fact be old or new. This means that subjects can make two types of correct responses and two types of incorrect responses.

Imagine a situation in which you are given a yes-no recognition test. The experimenter tells you that you will receive $1 for every *hit* (responding "old" when the item is old) and will be penalized only 1 cent for every *false alarm* (responding "old" when the item is new). You are likely to respond "old" on almost every trial simply to maximize the amount of money. Imagine the reverse situation, where you are penalized $1 for every

Table 9.1 Possible outcomes in a yes-no recognition test

		Subject's Response	
		Yes	No
Test Item	*Old*	Hit	Miss
	New	False alarm	Correct rejection

NOTE: The probability of a hit is 1 minus the probability of a miss, and the probability of a false alarm is 1 minus the probability of a correct rejection. Thus, the probability of a hit plus the probability of a miss equals 1.0. The probability of a false alarm plus the probability of a correct rejection equals 1.0. Usually, researchers report only hits and false alarms.

false alarm and rewarded with 1 cent for every hit. Now, you are likely to respond "new" on almost every trial to minimize your financial loss. Although extreme, these cases illustrate the large role that response bias can play in altering subjects' behavior, which is independent of their actual ability to tell whether an item is old or new (discrimination). Researchers have tried many different ways of correcting for guessing in the yes-no procedure, but the most common is to apply ideas from signal detection theory (Green & Swets, 1966).

Signal Detection Theory

Signal detection theory was initially developed to examine performance in perception experiments in which the subject's task was to detect the presence of a signal (a tone, for example). It is assumed that there is always some baseline neural activity, referred to as *noise*. Every so often, the experimenter plays a tone. This produces two possible situations: Either there is just noise, or there is *signal plus noise*. The task facing the subject is to distinguish between the two. The response given is based partly on the subject's true ability to discriminate the combination of signal plus noise from noise only, and partly on response bias.

A similar task confronts smoke detectors. There will always be some smoke particles in the air (perhaps from slightly burned toast or accumulated dust on a powerful lamp); the important question is whether this smoke is of sufficient concentration to indicate a fire of concern. The detector must be sensitive enough to detect smoke (detect that there are smoke particles), but it must also decide whether the smoke particles detected indicate an actual fire. Hopefully, the smoke detector will have a hit rate of 1.0, and as low a false alarm rate as possible.

Signal detection theory assumes that when you see a test item in a standard recognition test, it produces a certain amount of evidence that the item was in the study phase. Signal detection theory further assumes that this evidence is normally distributed. The normal distribution (often referred to as a bell curve) can be thought of as a frequency distribution. A frequency distribution is simply a way of plotting the number of times you observe a particular value. For example, say you flipped a coin 10 times and counted up how many heads you got and wrote this number down. Then, you flipped the coin another 10 times and counted up how many heads you got. Say you did this 1024 times. How many times did you get 5 heads out of 10? How many times did you get 9 heads out of 10? If you actually did this, you should find numbers similar to the ones in Figure 9.1, which are very close to those found in a normal distribution.

When applied to recognition, signal detection theory assumes that old and new items differ in how familiar they appear to the subject. The average new item will appear less familiar than the average old item because the old items have been processed more recently. If you were tested on all old items, and calculated the amount of evidence supporting the idea that each item was old, you could produce a graph similar to the one shown in Figure 9.1 with coins. If you also did this for all new items, you would end up with two distributions, represented in Figure 9.2.

Signal detection theory assumes that there are two normal distributions, one for old items and one for new items. On average, a new item will provide you with less evidence that it was an old item than will an old item, which is why the "old" distribution is to the

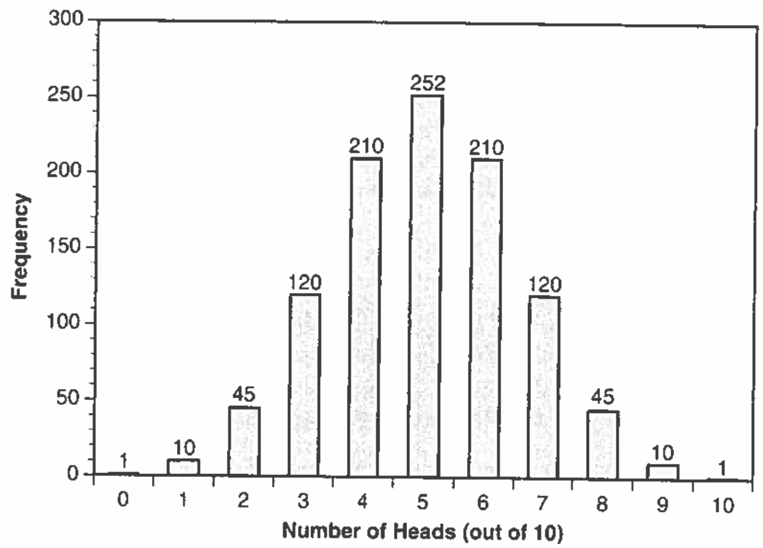

Figure 9.1 The expected frequency distribution if you tossed a fair coin 10 times, counted the number of heads, and then repeated this another 1023 times. You should find that you got only 2 heads out of 10 about 45 times, but 5 heads out of 10 about 252 times.

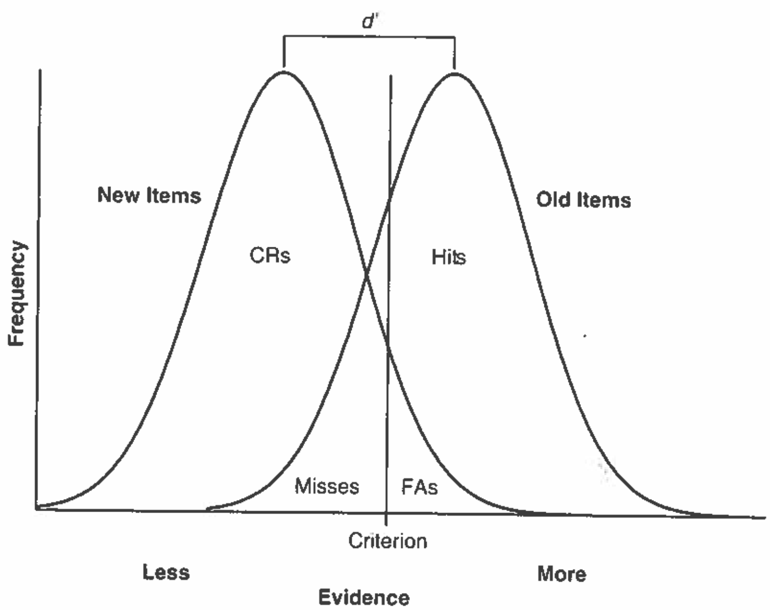

Figure 9.2 Signal detection theory as applied to yes-no recognition memory. The subject evaluates the evidence supporting the idea that the item is old or new. Old items are assumed to appear more familiar, on average, than new items, although there will often be some overlap. The distance between the means of these distributions is d', a measure of discriminability. The subject adopts a criterion, and items with more familiarity will be judged "old" (a yes response), whereas items with less familiarity will be judged "new" (a no response). Items that fall to the right of the criterion but are from the new item distribution are false alarms. Items that fall to the left of the criterion but are from the old item distribution are misses.

right of the "new" distribution. The difference between the means of these two distributions is called d' (pronounced "dee prime") and is the measure of bias-free discriminability. If old and new items appear equally familiar, then the difference between the means of the distributions will be 0. A d' of 0, then, represents the case where the distributions overlap and the subject cannot tell the difference. The greater the difference between the distributions, the larger d' is, and the more different the typical old and new items are.

How do you compute d'? It is simply the difference between the means of the two distributions divided by the standard deviation of the new distribution (noise). We cannot calculate these values directly; Figure 9.2 is a theoretical distribution. However, we can take advantage of some properties of the standard normal distribution. This is a distribution like the one shown in Figure 9.1, except that its mean is set to 0 and its standard deviation is set to 1. Any normal distribution can be converted into a standard normal distribution. In a standard normal distribution, a z score is simply the distance from the mean. So, a z score of 2 is two standard deviations above the mean; a z score of −1 is one standard deviation below the mean. Another property of the standard normal distribution is that the total area under the curve is 1.0. Half the area is above the mean, and half is below the mean. So, we find a z score such that the area above it equals the hit rate and call this z_{sn}. We also find a z score such that the area above it equals the false alarm rate and call this z_n. Then, $d' = z_n - z_{sn}$. Values of d' between 1 and 2 usually represent good yes-no recognition performance.

There are three main ways of obtaining the needed z scores. First, you can use tables that are included in most statistics textbooks. A more convenient way is to use a spreadsheet program. Most such programs have a built-in function that returns what is called "the inverse of the standard normal cumulative distribution." These functions take $1 - p(\text{H})$ and $1 - p(\text{FA})$ as arguments and will return z_{sn} and z_n. A third way is to use a special program. Many such programs are available on the Web, including one at http://rumpole.psych.purdue.edu/models/DPrimeCalculator.html.

If an experimenter uses a yes-no recognition test and reports only the proportion of items correctly recognized as old (hits), there is no way to assess the subjects' performance. Table 9.2 illustrates why this is the case. The first three rows all have hit rates of 90%, but d' ranges from 0 to 0.76 to 3.34. A d' of 0 means the subjects could not discriminate between old and new items, even though the hit rate is nearly perfect. Why is this the case? Because the false alarm rate is varying also. The last row shows how a hit rate of 50% can mean better discriminability than a hit rate of 90% when different false alarm

Table 9.2 Sample d' values for different hit and false alarm rates

$p(H)$	$p(FA)$	z_{sn}	z_n	d'	C
0.90	0.90	−1.282	−1.282	0.00	−1.282
0.90	0.70	−1.282	−0.524	0.76	−0.904
0.90	0.02	−1.282	2.054	3.34	0.384
0.50	0.02	0.000	2.054	2.05	1.029

NOTE: $p(\text{H})$ is the probability of a hit, $p(\text{FA})$ is the probability of a false alarm, z_{sn} is the z score such that the area above it equals the hit rate (sn denotes "signal + noise"), z_n is the z score such that the area above it equals the false alarm rate (n denotes "noise"), d' is $z_n - z_{sn}$, and C is $0.5(z_n + z_{sn})$.

Experiment *Recognition and Signal Detection Theory*

Purpose: To demonstrate the use of signal detection theory in analyzing yes-no recognition tests

Subjects: Thirty subjects are recommended; 10 should be assigned to the neutral condition, 10 to the conservative condition, and 10 to the liberal condition. Subject 1 should be in the neutral condition, Subject 2 in the conservative condition, Subject 3 in the liberal condition, Subject 4 in the neutral condition, and so on.

Materials: Table C in the Appendix contains a list of 96 two-syllable words randomly drawn from the Toronto word pool. For each subject, construct a list of 48 words in random order. The answer sheet should have all 96 words (in random order) followed by OLD and NEW.

Procedure: For each group, read to the subject the first set of instructions followed by the list of 48 words, at a rate of approximately 1 word every 2 seconds. At the end of the list, read the instructions appropriate for each group. Give subjects the prepared answer sheet, and have them circle either OLD or NEW for each item.

Instructions for All Groups: "I will read you a long list of words. After I have finished reading the words, I will give you a memory test. I will tell you more about the test after I have read the words."

Instructions for the Neutral Group: "On the answer sheet, you will see a list of 96 words. Half of these words came from the list I just read, and half are new words. I would like you to circle OLD or NEW beside each word to indicate if it was on the list. Because half of the words are new, if you are unsure of your response, it is no better to guess OLD than to guess NEW because each response is equally likely to be correct. Any questions?"

Instructions for the Conservative Group: "On the answer sheet, you will see a list of 96 words; 25% of these words came from the list I just read, and 75% are new words. I would like you to circle OLD or NEW beside each word to indicate if it was on the list. Because 75% of the words are new, if you are unsure of your response, it is better to guess NEW than to guess OLD because you will be more likely to be correct. Any questions?"

Instructions for the Liberal Group: On the answer sheet, you will see a list of 96 words; 75% of these words came from the list I just read, and 25% are new words. I would like you to circle OLD or NEW beside each word to indicate if it was on the list. Because 75% of the words are old, if you are unsure of your response, it is better to guess OLD than to guess NEW because you will be more likely to be correct. Any questions?"

Scoring and Analysis: For each list, count the number of times an old pair was judged to be old (hits) and the number of times a new pair was judged to be old (false alarms). The text describes three ways to calculate d' and C.

Optional Enhancements: Include more subjects in each condition, and collect a measure of confidence. For each item, after the old/new judgment, have the subject write down a number from 1 to 6 to indicate confidence. A 1 means very low confidence; a 6 means very high confidence.

SOURCE: Based on an experiment by Knoedler (1996).

rates are taken into account. The measure of response bias when applying signal detection theory to recognition memory is C (see Snodgrass & Corwin, 1988, for why C is preferred to b). C is the mean of the z_n and z_{sn} values. A value greater than 0 indicates a conservative response bias, a tendency to respond "new" more often than "old." A value less than 0 indicates a liberal bias—a tendency to respond "old" more often than "new." A well-analyzed study of yes-no recognition memory will present both the hit and the false alarm rates, a measure of discriminability (d'), and a measure of response bias (C). For hit or false alarm rates of 1 or 0, there is a standard correction so that d' and C can still be calculated (see Snodgrass & Corwin, 1988).

Signal detection theory is not the only way of analyzing data from yes-no recognition tasks. It assumes that the variance of the two distributions is the same, and it also assumes that the distributions are normal. A' is a nonparametric analog of d' (Pollack & Norman, 1964) that has been shown to be highly correlated with d' (Snodgrass, Volvovitz, & Walfish, 1972). It ranges from 0 to 1, with 0.5 reflecting chance performance. Because of the way it is calculated, it allows analysis of data from subjects who have hit or false alarm rates of 0, and it does not require the assumption of homogenous variance. Snodgrass, Levy-Berger, and Haydon (1985) show how to calculate A'. The appropriate measure of bias for A' is B''_D. This measure ranges from −1 to 1, with 0 indicating no bias; a positive number indicates a conservative bias (see Donaldson, 1992).

A third method is generally referred to as the *two-high-threshold model* (see Feenan & Snodgrass, 1990). The name comes from the idea that there is a threshold for old items and a threshold for new items (hence "two thresholds"), and only items that exceed the threshold will be recognized (hence "high"). The discrimination measure is *Pr*, also known as the corrected recognition score. It is simply the difference between the hit and false alarm rates. The bias measure, *Br*, is the false alarm rate divided by 1 minus *Pr*. A value of *Br* greater than 0.5 indicates a liberal response bias; a value less than 0.5 indicates a conservative response bias.

Feenan and Snodgrass (1990) recommend reporting not only hit and false alarm rates and d' and C (or A' and B''_D, if more appropriate), but also *Pr* and *Br*. The reason is that that some effects are observable in only a subset of these measures. For example, in their Experiment 1, subjects saw line drawings and were then given a recognition test in either the same context or a different context. The hit rate was larger when the pictures were tested in the same context compared to when they were tested in a different context, but there was no difference in the false alarm rates. Similarly, there was a difference in the *Pr* measure of discrimination but not in d'. Both measures of bias, however, revealed a large and statistically significant effect of changing the context: With a different context, subjects responded more conservatively than when the test context was the same as the study context. Although it is better to report a complete analysis, for many of the studies reported in this section, we will focus on only one measure to make the presentation more concise.

Single Process Models

Early theories of recognition were of a class known as single process models. For example, Yntema and Trask (1963) proposed a *tagging model* of recall and recognition in which each item is tagged when it occurs. To determine whether a word had been presented on a list, the subject examines the word in generic memory and looks to see whether there is a tag. The tag encodes not only the mere fact of presentation but also the relative time of occurrence. This view explains why people are very accurate at judging which of two words occurred first (Yntema & Trask, 1963), because all they need to do is examine the tags. It also explains why distractor words that are similar to target words are often incorrectly labeled "old" (Anisfeld & Knapp, 1968). For example, seeing the word *beach* might make the subject think of ocean, and *ocean* gets tagged along with beach.

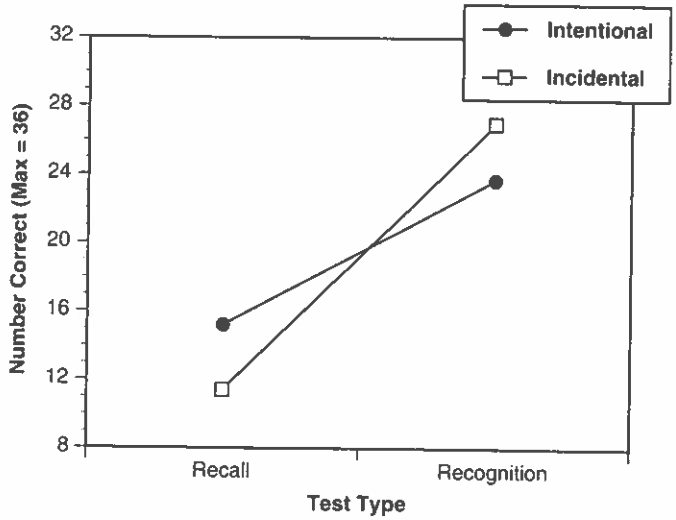

Figure 9.3 The interaction between learning instructions (intentional versus incidental) and test type (free recall or recognition).
SOURCE: Eagle & Leiter (1964).

A second type of single process model is *strength theory* (Bahrick, 1970; Wickelgren & Norman, 1966). The basic idea is that the more recently a particular item was experienced, the stronger or more familiar it seems. Strength could be used as the dimension in a signal detection type analysis.

The key limitation in both of these models is that they contain only a single process. With only a single mechanism, the same manipulation has to have the same effect regardless of the task; there is no provision for different processes. Evidence inconsistent with both of these views comes from studies that report different effects of the same manipulation depending on whether the test is recall or recognition.

Eagle and Leiter (1964) presented one group of subjects with a list of 36 words and told them they would have to remember the words. This was the intentional learning condition. Subjects recalled 15.2 words and recognized 23.7 words. In the incidental learning condition, a different group of subjects was given the same 36 words and told to classify the words based on part of speech. When a surprise memory test was given, the subjects recalled only 11.4 words but recognized 27.0 words. The key result, shown in Figure 9.3, is that recall is higher when intentional instructions are given than when incidental instructions are given (15.2 versus 11.4), but the opposite is true for recognition: Performance is better when incidental instructions are given than when intentional instructions are given (23.7 versus 27.0). Estes and Da Polito (1967) reported similar results. One reason may be that when intentional instructions are given, subjects can engage in appropriate strategies to organize the material, and organization has larger benefits for recall than for recognition (Kintsch, 1970; Mandler, 1967; see also Hunt & Einstein, 1981).

A second result of interest is the differential effect of word frequency. High-frequency words are recalled better than low-frequency words are, but low-frequency words are recognized more accurately than are high-frequency words (Deese, 1961; Gregg, 1976; Hall, 1954). Word frequency is normally expressed as the number of times the word is likely to

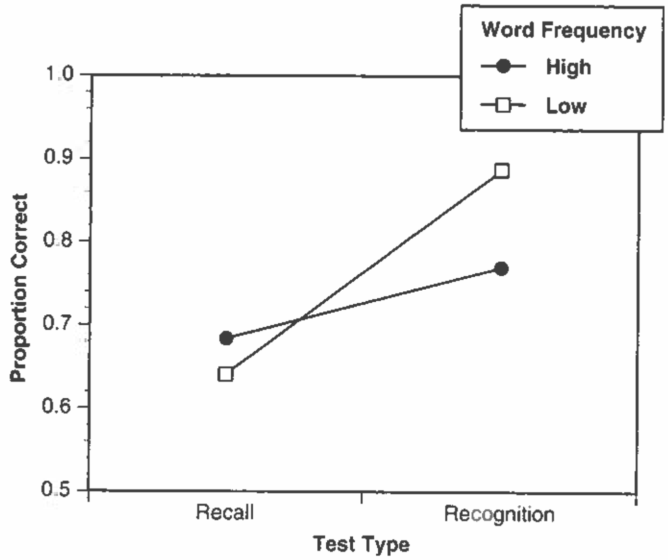

Figure 9.4 The word frequency effect in recall and recognition.
SOURCE: Based on data from Kinsbourne & George (1974).

be encountered per million words; it is computed by counting the number of occurrences in several different kinds of written documents (such as newspapers, novels, and magazines). Kinsbourne and George (1974) presented subjects with a 16-item list of words that were either of high frequency (words that occurred no fewer than 200 times per million) or of low frequency (words that occurred no more than 15 times per million). Half the subjects received a free recall test, and half received a recognition test; the results are shown in Figure 9.4. More high-frequency words than low-frequency words were recalled correctly, but more low-frequency words than high-frequency words were recognized.

Neither single process model can account for these results. With only a single process, a manipulation such as word frequency or intentionality must have the same effect on both recall and recognition. To overcome the limitations of single process models, researchers have developed a class of two-stage models known as generate-recognize models.

Generate-Recognize Models

Single process models were quickly replaced by a class of two-stage models collectively known as *generate-recognize models* (Anderson & Bower, 1972; Bahrick, 1970; Kintsch, 1970). According to these models, recall is made up of two processes, but recognition is made up of only one. In a free recall test, subjects must first generate a set of plausible candidates for recall. Once the set is generated, the subject then has to confirm whether each word is worthy of being recalled. (It is unfortunate that the second stage is called "recognition," which leads to confusion with a recognition test.) In a recognition test, the subject does not need the generation stage; the experimenter has provided the candidate. All that remains is the confirmation or recognition stage.

How would such a model explain the two findings described in the previous section? Anderson and Bower's (1973) version, called HAM *(human associative memory)*, begins with the assumption that words are stored in an associative network of nodes (see Chapter 10, e.g., Figure 10.5). Each node represents both a word and the concept represented by the word, and the nodes are connected by pathways to related nodes. As each word is presented, the node gets tagged with a contextual marker. Contextual markers contain information about salient stimuli—for example, a clock that was ticking, a door that slammed, or a siren in the background. If one word is associated with another word, the pathway can also be tagged. This is a little like Hansel and Gretel: As they went through the forest, they left bread crumbs to mark their passage, with the hope of retracing their steps. At recall, the subject follows the contextual markers to generate a set of plausible candidates.

The second stage, recognition, examines the number of associations between the target word and the context associated with the particular list. If there is sufficient contextual evidence, the subject is willing to say "old." If there is not sufficient evidence, the subject will say "new." Thus, the signal detection analysis shown in Figure 9.2 is again applicable: Recall will be enhanced to the extent that there is a rich network and lots of pathways have been tagged, and recognition will be enhanced to the extent that individual words are associated with particular contextual elements.

The intentional/incidental learning dissociation is easily explained. When learning is incidental, the subject does not associate words in the list with each other because there is no reason to do so; this will hurt recall during the generation stage. However, because the subject focuses entirely on one word at a time, there will be a strong association between the word and the contextual elements; this will help recognition. When learning is intentional, most subjects will adopt a strategy of associating each word in the list with other words. According to HAM, this will set up a richly marked network with lots of pathways tagged, facilitating generation and thus helping recall. At the same time, this strategy of associating words with one another will result in only weak associations between a given word and context, hurting recognition.

A similar analysis explains the word frequency effect. High-frequency words tend to have more associates and thus more pathways. Subjects should be able to find a shorter, more direct path between the nodes for a given list of short items. Low-frequency words, on the other hand, have fewer associates and can take longer to read. This makes it less likely that a short path can be obtained and so hurts recall. However, because many low-frequency words are unusual looking, they can take longer to process and thus lead to more item-context associations; this helps recognition. Indeed, the more unusual looking the word, the better the recognition of the item (Zechmeister, 1972).

There is one major problem with most versions of generate-recognize models: They require that if a word can be recalled, it must also be recognized (Watkins & Gardiner, 1979). Because the second stage is the stage that both recall and recognition have in common, a successful outcome at this stage in one test means a successful outcome at this stage for the other test. The reason a word can often be recognized when it is not recalled is due to the extra preliminary stage on a test of recall: Even though the item was not generated, it was capable of being recognized. For example, you may not be able to recall the name of the actor who starred in one of your favorite films, but as soon as someone suggests the name, you can accurately recognize it. Thus, recall failure of recognizable words is quite common. Tulving and his colleagues (Tulving & Thomson, 1973; Watkins & Tulving, 1975), however, have demonstrated a phenomenon known as *recognition failure*

Table 9.3 One procedure used to demonstrate recognition failure (Step 5b) of recallable (Step 6) words

Step	*Procedure*	*Example*
1a	List 1 presented	*badge—button*
1b	Cued recall of List 1	*badge—button*
2a	List 2 presented	*preach—rant*
2b	Cued recall of List 2	*preach—rant*
3	List 3 presented	*glue—chair*
4a	Free association stimuli presented	*table*
4b	Free association responses made	*table—chair, cloth, desk, dinner*
5a	Recognition test sheets presented	*desk top chair*
5b	Recognized items circled	*desk top chair*
6	Cued recall of list 3	*glue—chair*

SOURCE: Based on Watkins & Tulving (1975).

of recallable words. That is, contrary to the prediction of generate-recognize models, a word can be recalled under certain conditions even though it cannot be recognized.

The procedure used by Watkins and Tulving (1975) is shown in Table 9.3 (but simplified a little). The first two steps consist of a traditional paired-associate task, where the task is to recall the second word when given the first word as a cue. In Step 3, the critical list is presented, but note that it is not tested immediately. The important aspect of Step 3 is that the cue word is a weak associate of the target word (see Chapter 5). Step 4 is a free association test in which subjects are asked to generate as many associates as they can think of to a target word. Because the target word (*table*) is a strong associate of the target word in the previous step (*chair*), the subject usually provides the response term from Step 3. In Step 5, the subject is given a forced choice recognition test; this is where recognition failure can occur. The final step is the cued recall test for the list presented in Step 3. At this step, the subject often produces the word that was not recognized in the previous step. Watkins and Tulving (1975) found that 49% of the recalled items were not recognized. In Experiments 2 through 6, this value varied from a low of 16% to a high of 62%. They also determined, by using various techniques, that "the hocus pocus procedures of the early experiments turned out not to have been necessary to produce the phenomenon of recognition failure of recallable words" (p. 26). In particular, the free association step can be omitted and recognition failure can still occur.

Based on the encoding specificity idea presented in Chapter 5, the explanation of this phenomenon is quite straightforward. Contrary to the fundamental assumption of generate-recognize theory, recognition and recall both depend on the cues available at test. In the situation concocted by Watkins and Tulving, the cues available at test were better in the recall test than in the recognition test. (A more detailed account is presented in the optional chapter at the end of this book.) Recognition is not easier than recall, and recall is not easier than recognition; performance on each test depends on the cues available.

Beyond Simple Generate-Recognize Models

One change to a simple generate-recognize model that will allow it to account for recognition failure of recallable words is to have a search process occur during the recognition phase. For example, it is possible to have the same search and confirmation process operate in both recognition and recall (Jacoby & Hollingshead, 1990). The main problem with this approach is that subjects can very quickly and with a great deal of confidence correctly say that an item was not presented (Atkinson & Juola, 1974). Given the speed and confidence with which subjects reject an item, it seems unlikely that an extensive search of memory occurs every time.

Several researchers have suggested a process whereby recognition can use a search but can also rely on a simple familiarity process (Atkinson & Juola, 1973; Mandler, 1980). The idea is that a measure of familiarity is instantly computed. If this value is very large, then the subject gives a very rapid "old" response. If this value is very low, then the subject gives a very rapid "new" response. It is only for the intermediate familiarity values that a search takes place. A fundamental assumption is that the process of assessing familiarity is faster than the search process.

Mandler (1980) is careful to distinguish two types of recognition judgments: simple recognition, in which a judgment of prior occurrence is made, and identification. The former process may be accomplished solely by an evaluation of the item's familiarity, but the latter requires both familiarity and retrieval. Familiarity is related to the ease of processing the item: The more recently the item has been perceived, the easier it is to perceive the item at test. Both processes, familiarity and retrieval, are assumed to be initiated simultaneously. Although it has not been directly applied to the recognition failure of recallable words, Mandler's (1980) model has mechanisms that should allow it to produce the appropriate pattern. In the procedure detailed in Table 9.3, there is a relatively long delay between learning and the recognition test. This delay could lower the feelings of familiarity until they fall into the intermediate range, and performance must then rely more on the retrieval component. Because the target word in the recognition test is presented in a context different from the study context, the retrieval phase contains inappropriate cues. The recall test (Step 6) provides appropriate cues, so here the retrieval process will succeed. Note that this idea is similar to the encoding specificity idea presented in Chapter 5.

Mandler (1980) did report a simulation of a study of recognition that supports this analysis. Subjects studied word pairs and then received a variety of tests. For example, a subject studied the pair A-B; recognition was then assessed when just A was presented, when just B was presented, and when both A and B were presented. According to the analysis outlined above, performance will be better for the A-B items than for either A or B individually. This is exactly what Table 9.4 shows. In particular, notice that whereas both the hit rate and the false alarm rate improve when both A and B are presented at test, there is a larger difference in the false alarm rate. These results are nicely predicted by the model. Because the single A and B items are being tested in a new context, performance is worse than when tested in the old context (both items together).

Given the success on the recognition component, Mandler (1980) looked at recognition performance of B when B has been either recalled or not recalled (using data from Rabinowitz, Mandler, & Barsalou, 1977). According to the analysis presented above, the familiarity values for recognizing B should be relatively constant and in the intermediate range. This will cause more reliance on the retrieval component. When B was recalled, it

Table 9.4 Observed and predicted hit rate and false alarm rate for word pairs

Items	Hits		False Alarms	
	Observed	*Predicted*	*Observed*	*Predicted*
A	0.76	0.79	0.18	0.23
B	0.77	0.79	0.23	0.22
AB	0.86	0.81	0.06	0.04

SOURCE: Mandler (1980).

was recognized correctly with a probability of 0.69; when B was not recalled, it was recognized with a probability of 0.41. Despite this large difference, the estimates of the familiarity component were identical and were in the intermediate range: 0.40 and 0.39, respectively. A combination of familiarity and retrieval, then, can produce recognition failure of recallable words.

Gillund and Shiffrin (1984) conducted three experiments to test the assumption that both familiarity and a search process are important. The basic idea was to force subjects to respond very quickly or to make them wait a while. The average response time in the fast response condition was approximately 500 ms, compared to 2.5 to 3 s in the slow condition. In the fast condition, subjects should be relying more on the familiarity process, whereas in the slow condition, they should be relying more on the search process. Gillund and Shiffrin reported two main results. First, and not surprising, subjects were more accurate in the slow condition than in the fast condition. Second, and more important, there were no other differences between slow and fast responses. To the extent that the two processes are different, Gillund and Shiffrin argued, there should have been some interactions. Because they did not observe any interactions, they argued that a search process for recognition was not required and that recognition could be based solely on familiarity.

Current models of recognition are part of the so-called global memory models, such as SAM (Gillund & Shiffrin, 1984), MINERVA 2 (Hintzman, 1988), and TODAM (Murdock, 1982). These models, along with a connectionist model of recognition, are presented in the optional chapter at the end of this book. Interestingly, however, the type of view suggested by Mandler (1980) is making a comeback of sorts (see, for example, Reder et al., 2000). The mechanisms he suggested accord nicely with results from the remember/know paradigm that asks people to make distinctions more fine grained than simply whether they recognize the item.

Remember Versus Know

One relatively recent change in recognition methodology concerns the *remember/know* procedure (Gardiner, 1988; Tulving, 1985), although the ideas behind it were well articulated by Mandler (1980). In this procedure, subjects are given a recognition test and are asked to indicate whether they actually remember the information (have a conscious recollection of the information's occurrence on the study list) or just somehow know the answer (know that the item was on the list but have no conscious recollection of its actual occurrence). In Tulving's (1985) study, the first using remember/know methodology, subjects studied pairs of words in which the second word was a member of the category indicated by the first word, such as *musical instrument—viola*. The subjects then received a standard free recall test, followed by a cued recall test with the category name as a cue, and finally a cued recall test with the category name as a cue plus the first letter of the target item. The proportion of "remember" judgments decreased over the three kinds of tests and also decreased with increasing retention interval, relative to overall recognition performance.

Several other variables have been found to have different effects on remember compared to know responses; most of these have been reported by Gardiner and his colleagues. Gardiner (1988) found a levels of processing (Chapter 5) effect on remember judgments but not on know judgments, and also found a generation effect (Chapter 6) for remember judgments but not for know judgments. Gardiner and Java (1990) found the standard better recognition of low-frequency than high-frequency words for remember judgments but not for know judgments. Gardiner and Parkin (1990) had subjects engage in a secondary task during study and found that this divided attention manipulation disrupted remember judgments but not know judgments. Gardiner and Java (1991) have also shown that performance as assessed by remember judgments decreases more quickly over a 6-month period than does performance on know judgments.

CogLab Experiment
http://coglab.wadsworth.com/experiments/RememberKnow/

As an explanation, Gardiner (Gardiner & Parkin, 1990; Gardiner & Java, 1993) has suggested that remember judgments are influenced by conceptual and attentional factors, whereas know judgments are not; instead, they may be based on a procedural memory system, much like Schacter's (1994) PRS. This distinction sounds much like that between the factors that affect explicit memory and those that affect implicit memory (see Chapter 7).

Rajaram (1993) reported a series of experiments that examined remember/know judgments in more detail. Her first experiment replicated Gardiner's (1988) findings of a large level of processing effect on remember judgments, but no effect (if anything, a slight reversal) on know judgments (see the left panel of Figure 9.5). Her second experiment involved showing either pictures or words at study and only words at test. The task was to say whether either the word or the picture named had been seen previously. The standard picture superiority effect (Madigan, 1983) was observed for remember judgments: Performance was better for pictures than for words. However, there was again a reversal for know judgments: Performance was better for words than for pictures (see right panel of Figure 9.5). Although this finding is consistent with the idea that remember judgments are based on the explicit system and know judgments are based on implicit memory, it is also consistent with the idea that remember judgments may depend more on recollective processes whereas know judgments are based more on familiarity.

Rajaram's (1993) final experiments compared know judgments to judgments of confidence. In addition to distinguishing between remember and know judgments, Rajaram had subjects categorize their judgments in terms of whether they were sure or not sure. She again found a difference between remember and know judgments, but found no difference in confidence ratings. This important finding suggests that know judgments are not based solely on the perceived confidence of the subject. (However, see Donaldson, 1996, for an argument that remember/know differences might be attributable to a criterion shift.)

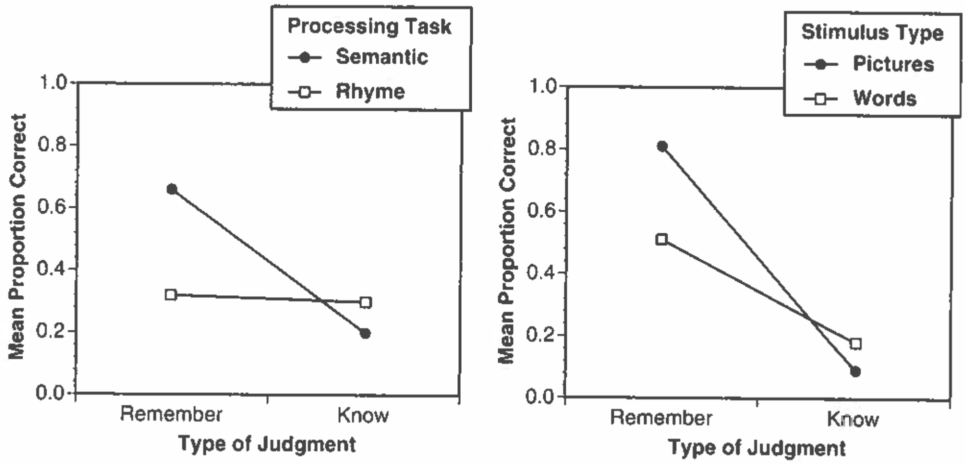

Figure 9.5 Two differences between remember and know judgments. The left panel shows the basic level of processing effect; the right panel shows the picture superiority effect, but only for remember judgments. SOURCE: Based on data from Rajaram (1993).

Recollection and Familiarity

The data from the remember/know paradigm illustrate quite nicely the idea that recognition performance is a combination of two different types of processes (Yonelinas, 2002). In one process, usually called *recollection*, memory judgments are made on the basis of conscious recollection of information about previous events. This process is similar to that used in free recall. In the second process, usually called *familiarity*, memory judgments are made on the basis of the assessed familiarity of a particular stimulus.

In addition to the remember/know paradigm, researchers have used a variety of tests to investigate recollection and familiarity. Yonelinas (2002) distinguishes between *task dissociation* and *process dissociation* methods. Task dissociation methods compare performance on two tasks that are thought to differ in their reliance on recollection and familiarity. For example, a response deadline task forces subjects to make a response at a particular time after the stimulus is presented. The subject might see the test stimulus and be asked to respond whether the item is old or new within 500 ms or within 1000 ms. The idea is that, because familiarity is supposed to be faster than recollection, fast responses should rely more on familiarity than recollection, whereas slow responses should include a larger contribution from recollection. Process dissociation methods generally compare performance in two conditions. In one (called the *inclusion* condition), both recollection and familiarity contribute to accurate performance, whereas in the second (the *exclusion* condition), the processes are set in opposition. Chapter 5 reviews this logic in detail.

Yonelinas (2002) reviewed empirical data from studies using both task dissociation and process dissociation techniques. Table 9.5 summarizes the effects of various encoding and retrieval manipulations on recollection and familiarity. Each row describes an experimental manipulation that occurred either at encoding or at retrieval and the effects of this manipulation on the two recognition processes. The table should be read in the following way: Relative to shallow processing (focusing on the perceptual characteristics of

Table 9.5 Summary of the effects of encoding manipulations and retrieval manipulations on recollection and familiarity

	Recollection	*Familiarity*
Encoding Manipulations		
Shallow vs. Deep Processing	Larger increase	Smaller increase
Read vs. Generate	Larger increase	Smaller increase
Full vs. Divided Attention	Larger decrease	Smaller decrease
Shorter vs. Longer Study Duration	Increase	Increase
Retrieval Manipulations		
Full vs. Divided Attention	Large decrease	No effect
Same vs. Different Perceptual Match (verbal)	No effect	Large decrease
Same vs. Different Perceptual Match (nonverbal)	Decrease	Decrease
No Delay vs. Short Delay	No effect	Rapid decrease
No Delay vs. Long Delay	Decrease	Decrease
Less Fluency vs. More Fluency	No effect	Large increase
Change Response Criterion	No effect	Large effect

NOTE: "Increase" and "decrease" refer to the effect on performance relative to the appropriate control condition.

a stimulus), deep processing (focusing on the meaning of a stimulus) led to an increase for both recollection and familiarity, but the increase was larger for recollection than for familiarity. The "larger increase" and "smaller increase" terms refer to the actual numerical differences calculated by Yonelinas for the experiments he reviewed. For the deep versus shallow manipulation, deep processing led to an increase of approximately 0.30 in the probability of recollecting the item, whereas deep processing led to an increase of slightly less than 0.20 for familiarity. The table summarizes this finding as a larger increase for recollection and a smaller increase for familiarity.

In general, encoding manipulations lead to similar effects on both recollection and familiarity, although the effects are usually larger on recollection than on familiarity. Retrieval manipulations, on the other hand, produce a pattern of results consistent with the idea that the two processes are independent. Familiarity processes are faster than recollective processes. Familiarity can be accurately modeled using signal detection theory, whereas recollection is better thought of as reflecting a threshold process. Both processes are sensitive to manipulations of conceptual information, although recollection shows larger effects than familiarity. Familiarity is more automatic than is recollection, although both can be affected by dividing attention at encoding.

Based on Yonelinas's (2002) review, it is clear that recognition memory is supported by at least two processes, one based on assessing the familiarity of the test item and the other based on recollection of information about previous events. This finding reinforces the conclusion from early studies, reviewed above, that ruled out single process models of recognition. A variety of dual process models have been proposed, including Mandler's (1980, 1991) model described earlier. Others have been proposed by Atkinson and Juola (1973, 1974), Jacoby (1991), Reder et al. (2000), and Yonelinas (2001). It remains unclear whether the so-called global memory models (SAM, MINERVA2, and TODAM, discussed in the optional chapter) can account for all of the results now attributed to recollection and familiarity. On the one hand, the global memory models are single factor models in that they assess a global measure of memory. On the other hand, different types of information can be included within that assessment. Thus, SAM can account for at least some of the process dissociation results using only the global assessment measure (Ratcliff, Van Zandt, & McKoon, 1995). The establishment of two processes within recognition may also shed some light on a problem that has long posed trouble for models of recognition: the mirror effect.

he Mirror Effect

One aspect of recognition that has received much attention in recent years is the mirror effect (Brown, Lewis, & Monk, 1977; Glanzer & Adams, 1985). In its simplest form, the *mirror effect* describes a regularity when examining performance on a variety of different kinds of recognition tests (such as yes-no, forced choice, multiple choice, and rating scales). A mirror effect is observed when "The type of stimulus that is accurately recognized as old when old is also accurately recognized as new when new. The type that is poorly recognized as old when old is also poorly recognized as new when new" (Glanzer & Adams, 1985, p. 8).

One example of a mirror effect concerns the effect of word frequency on recognition memory. Generally, low-frequency words are recognized more accurately than high-frequency words are (see Table 9.6 and Figure 9.4). However, word frequency also shows a mirror effect. Both of these effects can be seen in the data shown in Table 9.6. Rao and Proctor (1984) had subjects see a list of 120 words and then presented a recognition test of 240 words. The mirror effect can be seen clearly in their data: The number of hits was higher for the low-frequency words than for the high-frequency words; at the same time, the number of false alarms was lower for the low-frequency than for the high-frequency words. The standard word frequency effect can be seen in the larger d' value for the low- compared to the high-frequency words.

One reason this finding has attracted a great deal of attention is that the mirror effect seems pervasive. Glanzer and Adams (1985) reviewed all the published recognition studies they could find in which (1) a within-subjects design was used in which two (or more) levels of the variable were assessed and (2) the authors reported sufficient data (hit rates and false alarm rates for each stimulus class) to make the study useful. The results are shown in Table 9.7.

Table 9.6 The mirror effect (higher hit rates and lower false alarm rates in low-frequency than high-frequency words) and the word frequency effect in recognition

	Word Frequency	
	High	*Low*
Hits	27.84	31.00
False Alarms	10.20	7.63
d'	1.36	1.88

SOURCE: Based on data from Rao & Proctor (1984).

Table 9.7 The number of reports (separate studies or, in the case of the pictures versus words, separate experiments) that met the criteria and the number that showed a mirror effect

Variable	*Number of Reports*	*Number Showing a Mirror Effect*
Word Frequency	24	23
Concreteness	9	8
Meaningfulness	13	9
Pictures vs. Words	8	6
Miscellaneous	26	17
Total	80	63

SOURCE: From "The Mirror Effect in Recognition Memory," by M. Glanzer and J. K. Adams, 1985, *Memory & Cognition*, *13*, p. 8. Copyright © 1985. Reprinted by permission of Psychonomic Society Inc.

The one notable departure from the mirror effect comes from studies in the miscellaneous category that looked at recognition for normal as opposed to transformed text. With the transformed text studies removed, the number of mirror effects observed becomes 17 out of 22. Another possible exception is rare words (Wixted, 1992). Since the initial report, many studies have demonstrated mirror effects with other types of stimuli, including frequency discrimination (Greene & Thapar, 1994), associative information (Hockley, 1994), presentation rate (Ratcliff, Sheu, & Gronlund, 1992), age of the subject (Bäckman, 1991), and recency (Glanzer, Adams, & Iverson, 1991). It can also be seen with incidental learning (Glanzer & Adams, 1990) and when latency rather than accuracy is used as the response (Hockley, 1994). Even though they were excluded from the survey, experiments that use between-subject designs also show the mirror effect (Glanzer & Adams, 1985).

Having established the generality of the mirror effect, the next question is whether this effect is important or trivial. It could be that the mirror effect is an artifact of the way recognition data are analyzed. For example, a "mirror effect" occurs between false alarm rates and correct rejection rates. As false alarm rates increase, correct rejection rates decrease. This is an uninteresting mirror effect, however, because the false alarm rate and the correct rejection rate must add to 1.0 (see Table 9.1). Is this also the case for the mirror effect between hit rates and false alarm rates?

The consensus is that the mirror effect is not an artifact (Glanzer & Adams, 1985; Hintzman, Caulton, & Curran, 1994). One reason for this conclusion is that hit rates and false alarm rates can, in theory, vary independently; for any given hit rate, the false alarm rate can still be anywhere between 0.0 and 1.0. In the example given in the preceding paragraph, false alarm and correct rejection rates cannot vary independently; if the false alarm rate is 0.2, then the correct rejection rate must be 0.8. A second reason is that the actual pattern is not what is most likely. Let A represent a stimulus class in which old items are called "old" more often than B items, and also in which new items are called "new" more often than B items. To see a mirror effect, the distributions of these items needs to be ordered such that New A < New B < Old B < Old A. This is shown graphically in the top panel of Figure 9.6. Note, however, that this is not the only possible ordering; all old items could share a distribution, or all new items could share a distribution. In fact, given that

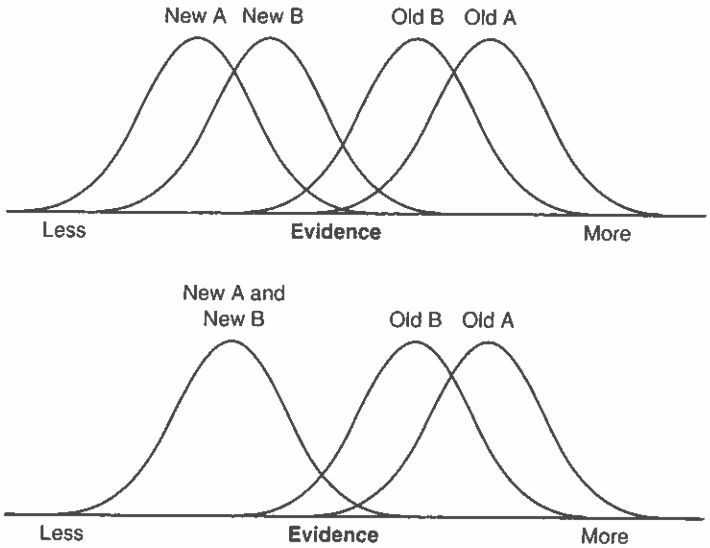

Figure 9.6 The top panel shows the required ordering of the four distributions of A and B stimuli according to signal detection theory if the mirror effect is to be seen. The bottom panel shows the most likely ordering of the distributions, where unstudied (new) items are unlikely to differ. SOURCE: Adapted from "The Mirror Effect in Recognition Memory," by M. Glanzer and J. K. Adams, 1985, *Memory & Cognition, 13*, p. 8. Copyright © 1985. Adapted by permission of Psychonomic Society Inc.

none of the new items has been studied, the most reasonable pattern is that the New A and New B distributions overlap, as shown in the bottom panel of Figure 9.6. However, the mirror effect is clearly the rule and not the exception.

Given that the mirror effect is both general and important, another important issue concerns its relationship to theories of recognition. The mirror effect clearly eliminates from contention all theories of recognition based on a unidimensional conception of strength or familiarity (Brown, Lewis, & Monk, 1977; Glanzer & Adams, 1985, 1990; Hintzman, Caulton, & Curran, 1994). It is unclear whether the global memory models (SAM, MINERVA 2, and TODAM, described in the optional chapter), which all rely on a unidimensional conception of familiarity, can explain the mirror effect. Although several new models of recognition have been proposed (Glanzer, Adams, Iverson, & Kim, 1993; Greene, 1996; Hintzman, Caulton, & Curran, 1994), it now seems likely that the mirror effect can be explained by dual process models (Balota, Burgess, Cortese, & Adams, 2002; Reder, Angstadt, Cary, Erickson, & Ayers, 2002).

For example, Joordens and Hockley (2000) have suggested that the mirror effect is due to two different factors, one of which affects the false alarm rate and the other the hit rate. The idea is that the mirror effect for word frequency occurs because high-frequency words are generally more familiar and low-frequency items are generally easier to recollect. With low-frequency words, the hit rate would be high because of better recollection, but the false alarm rate would be low because they are not very familiar. With high-frequency words, the hit rate would be lower because of reduced recollection, but the false alarm rate would be higher because they are more familiar.

Table 9.8 The mirror effect (higher hit rate and lower false alarm rate for low-frequency than high-frequency words)

	New Items (False Alarms)		*Old Items (Hits)*	
	High	*Low*	*High*	*Low*
Control	0. 26	0.18	0.75	0.80
Speeded Recognition	0.41	0.34	0.69	0.71

NOTE: The mirror effect is seen when unlimited time is allowed for the recognition judgment (the control condition). When speeded judgments are required, which should reduce the contribution of recollection, the hit rate advantage is eliminated but the false alarm advantage is unaffected.

SOURCE: Based on data from Joordens & Hockley (2000).

To test this idea, they conducted an experiment in which subjects were asked to recognized low- and high-frequency words. In the control condition, the subjects were given unlimited time to make their recognition judgments. In the speeded condition, they were told to make their response within 800 ms. The logic is that requiring a speeded response should reduce the influence of the recollection process but should not affect the familiarity process, because familiarity is faster than recollection. The results are shown in Table 9.8. With unlimited time, a standard mirror effect was observed. When speeded recognition was required, the hit rate advantage of the low-frequency items was eliminated but the false alarm rate advantage remained.

Currently, this account works well for many mirror effects that are produced by stimulus characteristics, but it does yet not offer a comprehensive account. A similar account has been formalized within a computational model (Reder et al., 2000). However, other explanations are probably needed for some forms of the mirror effect. Whatever the outcome, the mirror effect will play a large role in theory development in the future.

Face Recognition

Face recognition needs to be distinguished from face identification, a more difficult task. *Face identification* is being able to supply the name (and perhaps other details) that goes with a particular face, a form of paired-associate learning. *Face recognition*, in contrast, is deciding whether a particular face has been seen before. Bahrick (1984b) reported a study in which college teachers were asked both to identify and to recognize faces of former students. One variable of particular interest was the retention interval—the time since the end of the semester in which the student was in class. Bahrick's results are shown in Table 9.9. Face identification was consistently less accurate than face recognition.

Table 9.9 Percent correct recognition and identification of students' faces by their college teachers

	11 Days	*1 Year*	*4 Years*	*8 Years*
Recognition	69.0	47.5	31.0	26.0
Identification	35.5	6.0	2.5	0.0

SOURCE: Based on data from Bahrick (1984b).

It turns out that people make an extraordinary number of face identification and face recognition errors. Young, Hay, and Ellis (1985) had 22 subjects keep track of all of the errors they made recognizing and identifying people over an 8-week period. They reported 1008 errors, nearly 6 errors per day per subject. Of these, 114 were failures to recognize a person. Subjects rely heavily on physical features to recognize people—especially hair, forehead, eyes, and nose (Ellis, 1975)—and when these features change, recognition can fail. A second common error was mistakenly identifying one person as another; there were 314 such incidents, the most common being thinking an unfamiliar person is a familiar one. A third common error (233 reports) involved recognizing a person but failing to identify the person (as opposed to incorrectly identifying the person). These errors typically involved a slight acquaintance in a novel context, such as meeting your dentist in the grocery store. The fourth common error (190 incidents) was failing to recall the name of a person. In most cases, other information such as occupation was recalled. This finding, the so-called *name effect*, has been replicated: McWeeny, Young, Hay, and Ellis (1987) found that it took subjects longer to provide a name than an occupation. One nice control in this study was that the names and occupations were the same; for example, one face would be described as Mr. Baker, whereas a different face would be described as that of a baker. It was easier to say the occupation "baker" than the name "Baker."

Faces of people of the same race as the subject tend to be recognized slightly more accurately than faces of people from different races. For example, O'Toole, Deffenbacher, Valentin, and Abdi (1994) had a group of Caucasian subjects and a group of Asian subjects study Caucasian and Japanese faces. They performed a signal detection analysis on the recognition data, shown in Figure 9.7. Caucasian subjects recognized Caucasian faces more accurately than they did Japanese faces, and Asian subjects recognized Japanese faces more accurately than they did Caucasian faces. Similar conclusions have been reached by Valentine and Endo (1992) and Vokey and Read (1992).

Similar to the *other-race effect* just described is the *face inversion effect*. If even a very familiar face is shown upside down, the probability of correctly identifying or recognizing the face drops precipitously (Yin, 1969; Valentine, 1988). This face inversion effect has played a large role in the question of whether face recognition is a special process (see below). A face does not have to be shown rotated a full 180° to see a decrease in performance. A change in orientation of 45° (such as full face to three-quarter view) reliably impairs performance, and a change of 90° (full face to profile) results in even worse performance (Baddeley & Woodhead, 1983).

As with words and other stimuli, a form of priming can be observed with faces. If two well-known faces that are associated with each other are presented one after the other, the second face is identified more quickly than if it were an unfamiliar face (Bruce & Valentine, 1986). Thus, presenting a picture of Abbott will make recognition of Costello faster than if Abbott were not presented. Interestingly, if subjects believe that a face they are looking at belongs to a criminal, it is remembered more accurately (Honeck, 1986). The severity of the crime does not correlate with accuracy.

Two issues concerning face recognition and identification are how to explain the various phenomena and whether they involve a special process. Ellis and Young (1989) concluded that face recognition is special but not unique; faces are not the only stimuli to take advantage of it. As evidence that face recognition is special, they noted that (1) neonates show a preference for faces over nonfaces, (2) face recognition has a different developmental history than other phenomena, (3) faces are very difficult to identify upside

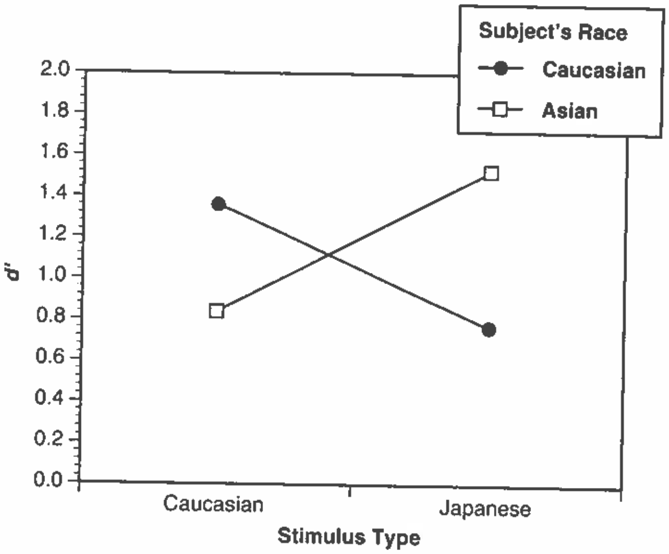

Figure 9.7 Accuracy (measured by *d'*) in recognizing Caucasian or Japanese faces as a function of the subject's race. SOURCE: Adapted from "Structural Aspects of Face Recognition and the Other-Race Effect," by A. J. O'Toole et al., 1994, *Memory & Cognition, 22,* 208–224. Copyright © 1994. Reprinted by permission of Psychonomic Society Inc.

down (the face inversion effect), and (4) there appears to be a special neural substrate that supports face recognition. Each of these factors will be discussed briefly.

The first line of evidence offered by Ellis and Young (1989) to support the conclusion that face recognition is special is the finding that newborn infants appear to display a preference for faces over nonfaces. However, the pattern of data is quite complicated. First, several studies show that face preference can disappear in infants at around 2 months of age, only to return a little later (see Maurer, 1985). Furthermore, other types of pictures show a similar privileged status; for example, infants may show a preference for pictures of normal over scrambled cars (Levine, 1989). Rather than arguing for specialness, this result could be seen as a general preference for coherence, or even for closed forms. A similar criticism has been offered regarding the second line of evidence, the special developmental history of face recognition. It turns out that this is not unique; both voice recognition and tonal memory have similar patterns of development (Carey & Diamond, 1980).

The third line of evidence concerns the face inversion effect. Diamond and Carey (1986) suggested that expertise—or, more accurately, a lack of expertise—may explain the face inversion effect. Subjects have a great deal of experience recognizing upright faces but almost no experience recognizing upside-down faces. Diamond and Carey used photographs of dogs as control stimuli for faces, and recruited experts in judging dogs. The dog experts showed a difference in recognizing upside-down versus upright dogs, just as people show a difference in recognizing inverted versus upright faces. The novices (people with little familiarity with particular dog breeds) did not show a difference. Similar explanations have been offered to explain the other-race effect.

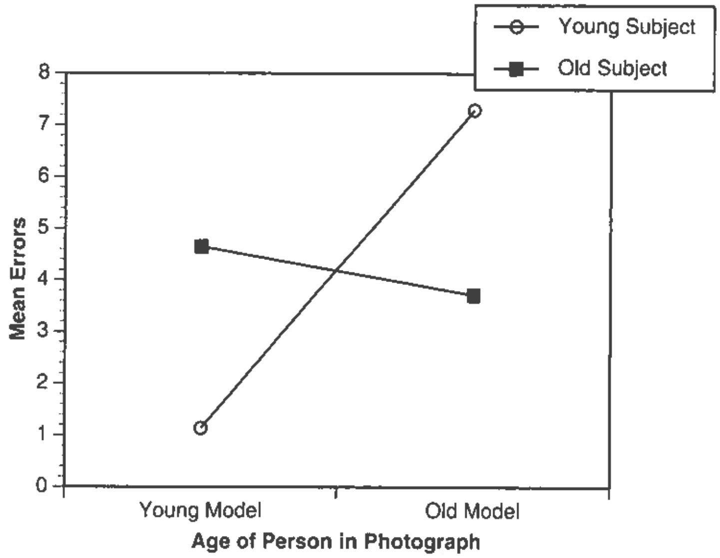

Figure 9.8 Mean errors observed in age estimates as a function of model age group and participant age group. SOURCE: Data from Winningham (2001).

This view predicts a phenomenon called the *other-age effect*. Assuming that young people interact more with other young people and that older people interact more with other older people, one should find that people process faces better if the faces are from their own age group. Winningham (2001) asked young (mean age 19.5 years) and old (mean age 77.3 years) subjects to estimate the age of models in photographs. Half the photographs were of young models (mean age 20.1 years), and half were of old models (mean age 79.9 years). As shown in Figure 9.8, the younger subjects were more accurate in estimating the age of the young models, and the older subjects were more accurate in estimating the age of the old models.

Valentine (1988) also views face recognition as more a matter of expertise than a special process. His conclusion comes from examining studies of inverted faces; rather than saying an upside-down face disrupts the special face recognition process, he argues that people have almost no experience with the task.

Ellis and Young's (1989) final point is that there appears to be some evidence of special neural substrates supporting face recognition. For example, the right hemisphere appears to be more involved in face recognition than the left, and several researchers have found cells in monkeys that respond only when the animal looks at faces. Furthermore, there is the phenomenon termed *prosopagnosia*, which refers to the inability to recognize familiar faces after certain brain injuries. As with the other points, this line of evidence is open to alternative explanations. For example, Levine, Banich, and Koch-Weser (1988) suggest that face recognition is not special, but is similar to other tasks that tap particular processes situated in the right hemisphere. In one study, they found a similar pattern of localized neural responding when the stimuli were houses. As another example, Dalesbred sheep have cells that respond to faces of sheep, dogs, and humans (Kendrick & Baldwin, 1987); however, at least one set of cells responded to familiar Dalesbred sheep rather than to unfamiliar Dalesbred sheep. A familiarity or expertise argument could again account for these findings. (Oddly, human and dog faces were responded to by the same cells, although the interpretation of this is not clear.)

One way in which processing of faces might differ from processing of other objects is that faces seem to be processed holistically. The *holistic encoding hypothesis* says that the encoded representation of the face is a unitary psychological entity (Farah, Wilson, Drain, & Tanaka, 1998). The key claim is that there is little or no decomposition of the encoded whole into its constituent parts. For example, subjects were presented with two faces simultaneously and were asked to judge whether one particular anatomical feature (such as the nose) was the same or different. Other features, irrelevant to the main task, could also be the same or different. Farah et al. found a substantial advantage when the irrelevant features were the same in the two faces, and this advantage was larger when the faces were upright than when they were inverted. This type of finding supports the idea of holistic representation because factors other than the target affect performance. It is as if the other anatomical features cannot be ignored. Wenger and Ingvalson (in press) suggest that the effect is due to decisional factors rather than to representing faces in a unitary way. In particular, it seemed as if subjects became more conservative in their judgment if irrelevant features varied but the relevant feature was same in the two faces.

Although there is still debate over whether face recognition is special, there is little evidence to support the argument that it is unique. Even such proponents as Ellis and Young think of it as only one of several special processes that have a similar set of properties. As more neurological evidence becomes available, there may be more consensus on the issue of how special face recognition is. In addition to advances in neurological study, analyses by several formal models of face recognition may also help (see, for example, Wenger & Townsend, 2001). For example, a recent connectionist model has simulated the other-race effect based solely on degree of experience with various kinds of faces (O'Toole et al., 1994).

Chapter Summary

In a standard recognition test, the stimulus can be old or new and the subject can respond yes, it is an old item, or no, it is not an old item. This results in four types of responses: hits (old item, yes), false alarms (new item, yes), misses (old item, no), and correct rejections (new item, no). Recognition performance cannot be determined unless both the hit rate and the false alarm rate are included. Several methods exist that try to determine how much of the response is due to knowledge (d', A', Pr) and how much to response bias (C, B''_D, Br).

Simple single process models of recall and recognition are ruled out by findings such as high-frequency words are recall better than low-frequency words but low-frequency words are recognized better than high-frequency words. The mirror effect also poses problems for these models. Generate-recognize models are ruled out by findings that under certain circumstances, words that cannot be recognized can be recalled. More information can be obtained by asking subjects to make remember or know judgments for all items they indicate are old. Findings from this paradigm are consistent with current views that recognition is made up of at least two processes, recollection and familiarity.

Face recognition (saying a face is familiar or not) is different from face identification (attaching a name to a face). Although people may be good at both, errors still occur. Face recognition does seem to be special but not unique; it seems closely related to expertise with processing the stimuli. For example, people are worse at processing unfamiliar kinds of faces (the face inversion effect, the other-race effect).

CHAPTER TEN

Knowledge

I consider that a man's brain originally is like a little empty attic, and you have to stock it with such furniture as you choose. . . . It is a mistake to think that the little room has elastic walls and can distend to any extent. Depend upon it there comes a time when for every addition of knowledge you forget something that you knew before.

—Sherlock Holmes

The more you put in a brain, the more it will hold.

—Nero Wolfe

Knowledge refers to what you know. You probably know the capital of Canada, the number of states in the United States, your telephone number, the number of legs on an ostrich, several types of dinosaurs, the color of carrots, the texture of sandpaper, the taste of seawater, and perhaps even the air-speed velocity of an unladen European swallow. Perhaps more surprising than the number of things you know is the number of things you know that you do not know. For example, you probably know that you do not know the name of the 44th element in the periodic table (molybdenum), the name of William McKinley's first vice president (Garret A. Hobart), or the name of the 1959 Nobel Prize winner for literature (Salvatore Quasimodo).

In this chapter, we look at various explanations for how people organize and retrieve information from semantic or generic memory. Although *semantic* memory is the most common term, a better term is Hintzman's (1978) *generic* memory, for three reasons: First,