SURENDRA N. SINGH
AND
GILBERT A. CHURCHILL, JR.

RESPONSE-BIAS-FREE RECOGNITION TESTS TO MEASURE ADVERTISING EFFECTS

URENDRA N. SINGH is sociate professor of marketing at the University of Kansas. He received his Ph.D. from the University of Wisconsin–Madison in 1982. His current research interests include development of response-bias-free recognition tests for measuring the memory effects of advertisements and examining the effect of programming-induced emotions and physiological arousal on the effectiveness of the commercials. His research has been published in the *Journal of Marketing Research*, the *Journal of Advertising*, and the proceedings of the APA, ACR, and AMA.

GILBERT A. CHURCHILL, JR. is the Donald C. Slichter Professor in Business Research at the University of Wisconsin–Madison. He was named Distinguished Marketing Educator by the American Marketing Association in 1986, only the second individual so honored. He is a former editor of the *Journal of Marketing Research* and has served on the editorial boards of the *Journal of Marketing Research*, the *Journal of Marketing*, the *Journal of Business Research*, and the *Journal of Health Care Marketing*.

Since its origination in the Second World War as a mechanism for making ground observations of enemy planes more accurate, the theory of signal detection (TSD) has been used to address a variety of problems. It has been used in such diverse areas as electrical engineering to aid in the design of sensing devices (Peterson, Birdsall, and Fox, 1954; Van Meter and Middleton, 1954) and statistical decision theory (Wald, 1950).

Psychologists, though, have particularly used TSD quite heavily. They have used it in the study of sensory-evoked potential (Hillyard *et al.*, 1971); speech perception (Egan and Clarke, 1956); animal learning (Rilling and McDiarmid, 1965; Suboski, 1967); memory (Parks, 1966; Bernbach, 1967; Hopkins and Schultz, 1969); audiology (Campbell and Moulin, 1968); attention (Moray, 1970; Sorkin *et al.*, 1972); and clinical psychology (Sutton, 1972). In fact, the theory can be applied to any situation where sensory input is ambiguous.

One area where TSD seems to hold great promise is in testing the recognition memory for advertisements. While psychologists have used TSD extensively in recognition testing (Banks, 1970), marketing practitioners and theoreticians have largely ignored it (Singh and Churchill, 1986). This article demonstrates how TSD can be used in recognition testing.

Advertising Recognition Testing and the Problems Associated with It

Recognition tests are extremely popular in measuring the memory effectiveness of print ads. In a typical recognition test, subjects are shown a series of ads, one at a time, and are asked to indicate whether they think they have seen it before.

Despite their popularity, recognition tests are often criticized because of their failure to account for respondent error. For example, Appel and Blum (1961) and Marder and David (1961) pointed out long ago that a large percentage of respondents will claim recognition of bogus ads (ads respondents could never have seen before) contained in magazines when real ads are also being tested. In some studies, the claimed level of recognition for bogus ads has been almost as high as that for real ads. Simmons (1961), for example, found that 32.4 percent of the respondents claimed to have seen an ad that did *not* appear in a two- to six-week-old issue of a magazine that they had read compared to 33.4 percent of the respondents who claimed reader-

---

This project was supported in part by the University of Kansas General Research Fund allocation #3401-20-0038. Support was also provided by the University of Kansas School of Business Research Fund. The ideas and opinions expressed herein are solely those of the authors.

ship for the ad that actually did appear. The general tendency for claimed level of recognition for bogus ads to be almost as high as that for real ads led to the conclusion in one study that

> . . . considerable general inflation is indicated by the fact that people who could not possibly have seen particular advertisements report varying amounts of "recognition." The practical range of "false" recognition may run from 5 to as much as 50 percent (Lucas and Britt, 1963).

One source of respondent error is acquiescence response-set bias. Response-set bias refers to the general influence of people's mental states on how they react to items, apart from their content. Acquiescence response-set bias refers to people's general tendencies to favor "yes" responses over "no" responses (Wells, 1961). Other sources of respondent error are guessing when uncertain; eagerness to please the interviewer; hesitation to appear ignorant; guessing on the basis of general familiarity with, or interest in, the advertised product or service and its advertising in general; freedom of the respondents to claim anything they wish; and the tendency of people to deny socially undesirable traits and to admit to socially desirable ones (Lucas and Britt, 1963; Clancy *et al.*, 1979).

A number of methods have been suggested over the years for making ad recognition tests more valid. Most of these methods rely on the use of "false" (bogus/distractor) ads along with "true" (stimulus) ads in the test. The responses of subjects on the distractor ads are then used to provide better estimates of their "true recognition memory."

The use of distractors actually serves two purposes. First, it makes respondents aware of the fact that they cannot indiscriminately claim recognition of items and this motivates them to pay greater attention to the task. Second, the responses to the distractor items provide a mechanism for adjusting the obtained scores to allow for respondent error. The methods vary primarily in terms of how the distractors are used to make the adjustment.

Early History. Lucas (1942) was the first to present a method for obtaining adjusted recognition scores—adjusted for guessing and other response biases—for magazine ads. His technique required the use of two samples of subjects. Subjects in sample 1 (control group) were shown the cover and some editorial features of the magazine issue in question. Those who qualified as readers were then shown the ads from the prepublication issue of the magazine along with an equal number of published (or familiar) ads bound in a portfolio. Subjects in the second sample were tested after the publication of the issue with a portfolio that included ads from the published issue along with an equal number of unfamiliar ads from the next prepublication issue. The logic was that subjects in the control group could not possibly have seen the ads since they were not yet published and their claims to the contrary were definitely false, whereas the readership claims made by the second sample (after publication) include both the effects of false claiming as well as actual exposure. An adjusted recognition score could thus be obtained by ascertaining the difference between post- and pre-publication scores through the equation:

$$\text{Adjusted audience} = 100 \times \frac{\text{Posttest score} - \text{Pretest score}}{100 - \text{Pretest score}}$$

Simmons (1961) also suggested that the use of two samples of subjects could provide a valid measure of recognition memory when one was exposed to the issue containing the ad and the other was not. He based his calculation of an adjusted score on the questionable assumption that those saying "yes" to an unexposed ad are equally as unreliable as those saying "yes" to a previously exposed ad.

Appel and Blum (1961) also argued for the use of matched samples of subjects. In fact, they interviewed two matched samples of *Life* readers. Sixteen bogus ads were inserted into a test issue of *Life*. One sample was previously exposed to the test issue (the pre-exposure group) whereas the other sample was not (the post-exposure group). The authors found a high correlation between the aggregate recognition scores for each ad across the two groups which caused them to regress the post-exposure scores on the pre-exposure scores for the test ads and to use the deviations of the exposed readers' scores from the predicted scores as a measure of recognition memory due to ad exposure.

There are several serious problems with these dual-sample approaches to measure ad recognition which help to explain why they never became very popular among advertising researchers. In the first place, they are inherently expensive in that they require fairly large samples to form both test and control groups. Second, since the adjusted recognition scores are based on two sample scores, they are inherently less stable than either of the individual sample scores and can produce nonsense results. For example, a high proportion of false claiming sometimes results in negative recognition scores using Lucas' (1942) method, a logical impossibility.

Not all of the early attempts to measure ad recognition used control group samples. Moran (1951a), for example, proposed that instead of showing respondents one advertisement in a recognition test that they be presented with two ads simultaneously in order to remove their response biases: "The respondent is then placed in the position of choosing between two advertisements and cannot, therefore, 'please' the interviewer simply by saying 'yes'." Moran's procedure was later challenged by Heller (1951) and defended by Moran (1951b). One of the main problems with his procedure—how to select the ads to be paired with each test-advertisement—still remains unsolved.

Davenport *et al.* (1961) suggested a method for obtaining true readership of newspaper ads that also did not rely on a control group of nonreaders. His procedure involved interviewing respondents using two separate measures of memory, a see-scale and a read-scale, respectively. Respondents were exposed to both true and false ads. Those respondents who claimed recognition of an ad on one scale but not the other were considered as nonreaders of that ad. A reader credibility score was computed for each subject by dividing the number of true ad claims by the total number of (true plus false) ad claims by the subject. Subjects scoring above the median credibility score were characterized as high credibles (HCs) and those below the median as low credibles (LCs). The percentage of HCs who claimed readership for a given ad was then taken as the true readership for that ad. Not only does Davenport's method suffer from the response bias introduced by having subjects complete the two scales immediately after each other, but the method wastes much of the data collected in that estimated readership for a given ad is based solely on the responses of the HCs.

More Recent History. Given the difficulties encountered in finding a way to correct for the response tendencies of subjects, the search for better procedures to assess recognition memory languished in the advertising literature for over a decade. At the same time, recognition remained a very popular measure of memory among practitioners, especially for magazine ads (Clancy *et al.*, 1979). In recent years though, the search for better measures of recognition memory has been rekindled, partially it seems because of the development of a parallel need to measure the effectiveness of broadcast ads. Krugman (1977), for example, persuasively argued that recognition, rather than recall, may be the most appropriate test of memory for ads dealing with low-involvement products advertised on television.

Zielske (1982) found that, as a copy-testing procedure, day-after recall understated the true remembrance of the "feeling" ads for television commercials but not for the magazine ads. On the basis of these findings, Zielske advocated the use of recognition instead of recall for testing the memory of television commercials when the purpose of such a test was to compare the performance of "thinking" and "feeling" commercials. At least one research firm, Bruzzone Research Company (BRC), has begun to test memory for television commercials using recognition measures.

Considering the popularity of recognition measures and in view of the very high false-recognition scores obtained for bogus ads in the early studies, Clancy *et al.* (1979) decided to investigate the propositions put forth by some proponents of recognition measures (e.g., Neu, 1961) that overclaiming for bogus ads was a result of respondents not being informed about the presence of such ads and that interviewers were not carefully trained and controlled. Informing respondents about the presence of bogus ads and using trained interviewers did not eliminate overclaiming of readership though, causing Clancy *et al.* (1979) to comment:

. . . the search for better measures of recognition memory has been rekindled, . . .

> If indeed a recognition measure of advertising effects under circumstances of low involvement is more in keeping with the brain's function, advertisers will face even more of a need to improve upon the current measurement errors associated with the recognition approach.

Two major studies dealing explicitly with recognition measures have appeared in the advertising literature since the above exhortation. Bagozzi and Silk (1983) were primarily interested in the theoretical issue of whether recognition and recall measured common or distinct memory processes. They found that recall and recognition did not measure a single memory state but that memory was multidimensional, and recall and recognition measured only a portion of it. However, when interest in the ads was held constant, recall and recognition did measure memory as a unidimensional construct.

Unfortunately, the recognition (as well as recall) data used in the above study came from the data collected by the Printed Advertising Rating Methods (PARM) committee in 1956 for the Advertising Research Foundation. The recognition data was obtained for 95 print ads contained in a single issue of *Life* magazine using an ordinary yes/no-type recognition test. Therefore, there is reason to believe that the data suffered from all the response biases that are inherent in such recognition testing procedures, and the findings about the dimensionality of memory processes based on such data are suspect.

Unlike Bagozzi and Silk (1983), who were interested in understanding the memory processes involved in recognition and recall, Singh and Rothschild (1983) explicitly investigated the effect forced-choice recognition tests have on recognition scores. A forced-choice recognition procedure is different from the usual yes/no-type recognition test. In a forced-choice recognition test, subjects are presented with two or more advertisements at a time. The task is to pick out the original (stimulus) ad from the distractor ad or ads. If the subject sees two advertisements at a time, then the test is called a two-alternative forced-choice—if three ads, a three-alternative forced-choice, and so on (Klatzky, 1980). Presumably, the overall tendency to say "yes" to an item should affect both alternatives (stimulus as well as distractors) alike and thereby should exert little influence on the observed choice (Shepard and Chang, 1963).

Using nine-alternative and five-alternative forced-choice tests, Singh and Rothschild showed that recognition scores differed significantly across the number of repetitions (1, 2, and 4) and across length of commercial (30 seconds versus 10 seconds). Also, they found that an increase in the number of distractors in a forced-choice recognition test lowered the recognition scores in general. In sum, they found that recognition scores derived from a reasonably difficult forced-choice process behaved as theory would suggest they should.

The forced-choice recognition testing procedure proposed by Singh and Rothschild (1983) may indeed be a viable option for removing the influence of noting set. Certain caveats with the method must be noted though. For one thing, recognition scores are sensitive to the number of distractors used in a forced-choice recognition test (e.g., Davis, Sutherland, and Judd, 1961; Murdock, 1963; Postman, 1950; Slamecka, 1967; Teghtsoonian and Teghtsoonian, 1970). The scores are also affected by the similarity of the distractors to the stimuli (e.g., Bruce and Cofer, 1967; Dale and Baddeley, 1962; Underwood, 1965; Deese, 1963). As the number of distractors increases, the recognition task becomes more difficult. Likewise, the more similar the distractor ads are to the stimulus ads, the harder is the recognition task and the lower are the recognition scores.

The evidence thus seems to suggest that, after 40 years of research, there is still the need for a procedure that controls for the response biases and discrimination abilities of subjects in testing their ad recognition, but a procedure which is not affected by the choice of the number or the similarity of the distractors that are used to assess their response biases. The theory of signal detection (TSD) seems promising in this regard.

Table 1
Four Possible Outcomes in a Signal Detection Task

	Signal was	
Observer says	Present	Absent
Yes	Hit	False alarm
No	Miss	Correct rejection

Theory of Signal Detection

The key notions in TSD can be understood from its application to recognition testing of memory. In a typical recognition memory test, the subject is presented with a list of items (e.g., words), a portion of which the subject has been exposed to in an earlier session while another portion are distractors. As each item is presented, one at a time, subjects are to respond "yes" if they think that the item was on the original list and "no" if it was not. Subjects are told beforehand the *proportion* of the items that are old (i.e., were on the original list) and that are distractors. Items to which subjects have been previously exposed should be familiar to them; in signal detection language, old or familiar items are called signals or stimulus items while new or distractor items are called noise (Banks, 1970). Subjects can be paid for every correct response and can be penalized for every incorrect response, typically by withholding the reward. Usually, there are also nonmonetary rewards (e.g., eagerness to please the interviewer, hesitation to appear ignorant, and so on) operating which affect the answers given by the subject.

There are four possible outcomes to every recognition trial—the subject either may say that the word was old or new and the trial may have been signal or noise. Table 1 depicts the possibilities. A *hit* response is one in which the subject says "yes" to the presence of a signal, and the signal was actually present; a *miss* occurs when the subject says "no," but the signal was present; a *false alarm* occurs when the subject reports the presence of a signal, but in reality the trial contained noise alone; finally, a *correct rejection* occurs when the subject says no signal was present, and the trial actually did not have a signal. Notice that the two cells in each column of Table 1 are dependent; that is, if we know the probability of one cell in each column, we can estimate the probability of the remaining cell. This is because the sum of the probabilities of "hit" and "miss" must equal 1.00; similarly, the sum of the probabilities of "false alarm" and "correct rejection" must add to 1.00. Alternatively, the probability of total "yes" and "no" responses given the signal was or was not present must sum to 1.00. Given this complementary relationship between the cells in each column, the two cells customarily used to describe the 2 × 2 response matrix are the "hit" and "false alarm" ratios.

The performance of a subject in the recognition test depends mainly on two factors: the ability of the subject to perform the task and the motivational state and response tendencies of the subject (Pastore and Scheirer, 1974). The experimenter can affect the subject's response tendencies and motivations by changing payoffs and/or by changing prior odds. For example, in a word recognition test, if the subject is aware that there is no penalty for incorrect answers, the subject would probably have a greater motivation for guessing than if wrong answers were scored negatively. However, the subject's discrimination ability should remain unaffected by changes in motivational factors. Unfortunately, these two aspects—the sensory or discrimination capabilities of subjects and their decision-making styles (e.g., the affect of their values, motivations, knowledge of prior odds, and so on)—are completely confounded in the responses that are secured from the subjects. *The basic aim and unique contribution of the TSD is the separation of the sensory capabilities of the subject from the individual's decision-making aspects and the precise estimation of each* (Coombs, Dawes, and Tversky, 1970).

A method is needed by which recognition scores for individual ads can be adjusted to account for the sample of subjects claiming to have seen them.

TSD makes a number of assumptions. The first is that any information that an individual possesses has a certain strength in long-term memory. The strength of the item can be taken as the strength of a memory trace or the degree of familiarity. The more familiar an item is, the greater would be the memory strength for it, and vice versa. The second assumption is that measurements of the strength of items, both old and new, are normally distributed and have equal variances. This means in essence that there are two normal distributions for subjects to consider, one representing the list of familiar items and one representing the list of distractor items. Finally, TSD assumes that an individual's exposure to an item increases its strength in the long-term memory of the subject. In other words, both the stimulus and distractor items (ads) have certain strength values to begin with, but the strength value is changed with exposure to the item during the experiment. This moves the distribution of stimulus items to the right of the distribution of distractor items although the two distributions can overlap.

These assumptions imply the following with respect to ad recognition testing where the essential question being asked is "Which of the following ads have you seen in the publication issue being evaluated?" Suppose, for the sake of argument, that one half of the ads are stimuli (i.e., come from the publication issue being evaluated) whereas the other half are distractor ads. Now, some of the distractor ads might have been very familiar to the subject previously, some might have been very unfamiliar while still others might have been moderately familiar. The assumption in TSD is that the distribution of these strength values for both the stimulus and distractor ads is normal. Since the subject is presumably exposed to the stimulus ads (but not the distractor ads) before the test though, their strength value increases compared to distractor ads which remain at their initial strength. In effect, the distribution of old ads on the familiarity continuum is moved to the right by a fixed amount. (For a detailed rationale supporting these assumptions, see Klatzky [1980] for the general argument and Singh and Churchill [1986] for its applicability to ad recognition tests.)

Based on these assumptions, TSD derives two parameters, β and d'. β is the measure of response bias. As a threshold for saying "yes," it is a function of the individual's response tendencies which may depend on a number of factors including the individual's motivations, attitudes, and the prior probabilities of the occurrence of the stimulus items in a given test. On the other hand, *d'* is a measure of the subject's recognition memory sensitivity or how well the subject is able to discriminate between stimulus and distractor items. Both β and *d'* can be computed by knowing only a subject's hit and false-alarm rates in a given recognition test. Published tables of β and *d'* are also available (see, for example, Elliott, 1964).

Under the assumptions that the two distributions are normal and have equal variance, β and *d'* turn out to be independent of each other, i.e., a change in β will not affect *d'* if there is no real change in the memory capability of the individual.

> It is a fact of fundamental importance to SDT (*signal detection theory*) that the two values, P(Y/o) (*or hit rate*) and P(Y/n) (*or false alarm rate*) covary when, all else being equal, motivation varies. If, for example, a subject has marked on a recognition list only those items he is sure are old, he will have a low H rate and a very low FA rate. If he is then asked to find some more old items, he must accept some items of which he is less confident, and will surely increase his FA rate as well as his H rate (Banks, 1970).

There are two major stumbling blocks to using TSD to improve ad recognition testing. The first is the strict Gaussian assumptions the theory makes regarding the shape of the distributions relating the memory strengths of the stimulus and distractor items in that the empirical evidence in the advertising context indicates that the assumptions dealing with both the normality of the distributions and the equality of their variances are frequently violated. Under such conditions it is preferable to use TSD-based distribution-free statistics (Green and Swets, 1966) to assess the discrimination abilities and response biases of subjects. Fortunately, these statistics too can be computed knowing only a subject's hit and false-alarm rates, although the question of their independence then becomes an empirical matter. The second problem is that the response bias and recognition memory parameters apply to subjects, not ads. A method is needed by which recognition scores for individual ads can be adjusted to account for the sample of subjects claiming to have seen them.

TSD-based Nonparametric Indices of Response Bias and Sensitivity

While a number of nonparametric indices are available (see Green, 1964; Green and Moses, 1966), most fail to preserve the distinction between sensitivity and bias. (For a detailed criticism of various measures see Grier, 1971.) The two nonparametric indices that are directly comparable to β and *d'* are B'_H (*B*-prime *H*) and *A'* (*A*-prime), respectively.

B'_H: A Nonparametric Index of Response Bias in Recognition Tests. The B'_H measure proposed by Hodos (1970) is based on the geometry of the unit square which is a square in which each arm is one unit long and which is obtained by plotting the hit rate as a function of the false-alarm rate as shown in Figure 1.

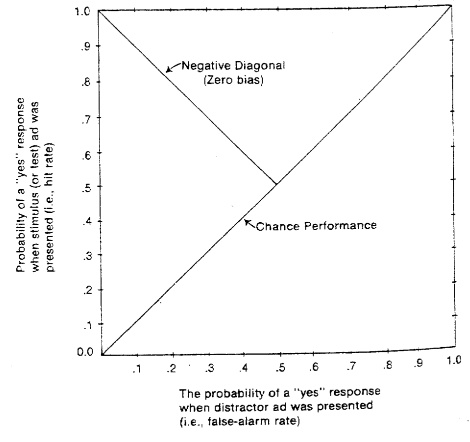
The positive diagonal line represents chance performance. For example, if a subject said "yes" randomly 60 percent of the time, the subject would have a hit rate of 60 percent when the test (stimulus) ads were actually present, but the subject's false-alarm rate would also equal 60 percent when the distractor ads were present. Points below the diagonal would occur only due to chance or when receivers are deliberately saying "yes" when they think they should have said "no" and vice-versa.

Assuming that subjects are extremely liberal in their responses and always says "yes," then they would be correct on every response when the stimulus ad was present but incorrect on every response when the distractor ad was present. Thus, they would have a hit rate of 100 percent but also a false-alarm rate of 100 percent. Their responses would then produce a point in the upper right-hand corner of Figure 1. Similarly, if subjects were extremely conservative and always said "no," the probability of their hit and false-alarm rates would equal zero and would produce a point in the lower left-hand corner of the unit square. Points on the negative diagonal going from the upper left-hand to the lower right-hand corner represent unbiased performance in that the subject is equally likely to respond "yes" or "no" under ambiguous stimulus conditions (Hodos, 1970).

A point falling to the right of

Figure 1

Unit Square



the negative diagonal represents the tendency of subjects to say "yes," whereas a point falling to the left of the diagonal represents the subject's tendency to say "no." Hodos used these fundamental ideas to develop a method that corrects for both "yea-saying" and "nay-saying" tendencies in recognition tests.

Given the coordinates (x,y) of a point p, to the left of the negative diagonal representing a subject's actual performance in a recognition test, the bias index B'_H can be computed using the formula (Grier, 1971):

$$B'_H = 1 - \frac{x(1-x)}{y(1-y)} \qquad (1)$$

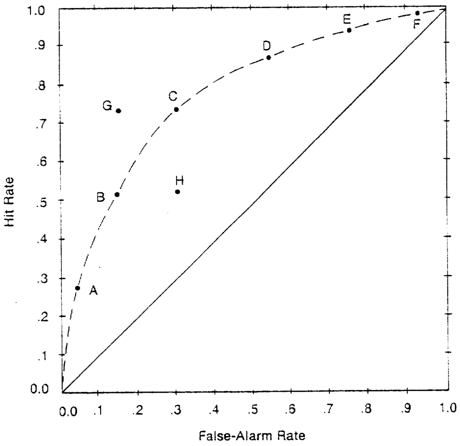
where y is the hit rate and x is the false-alarm rate. For points to the right, the formula becomes

$$B'_H = \frac{y(1-y)}{x(1-x)} - 1 \qquad (1)$$

According to this scheme then, a tendency to say "yes" is represented by a negative-bias-correction factor whereas a tendency to say "no" is captured by a positive-bias-correction factor, since the percent bias measure of the yea-saying tendency has a negative sign whereas the percent bias measure of the nay-saying tendency has a positive sign. This makes intuitive sense because a person with a tendency to say "yes" spuriously inflates recognition scores; therefore, the subject's score should be adjusted downwards and hence the negative sign when calculating B'_H. The reverse is true for "nay-sayers." As the formulas indicate, B'_H scores can vary from −100% to +100% representing maximum yea-saying and nay-saying, respectively.

Figure 2

The Memory-Operating Characteristic (MOC)*



*Adapted from Banks (1970)

A': A Nonparametric Index of Memory Sensitivity in the Recognition Tests. A' was proposed by Pollack and Norman (1964) as a measure of memory sensitivity. Like B'_H, it depends on the geometry of the unit square; however, it also uses the concept of a memory-operating characteristic (MOC) curve or simply MOC. A MOC is the graph of hit and false-alarm rates of a subject under *varying motivational states but the same memory conditions.* Assume Figure 2, for example, represents the MOC curve for a subject in a word-recognition test. Point A indicates that subject is acting very cautiously and saying "yes" to old items only when very sure of them. The subject has a low hit rate and a low false-alarm rate. If the subject can be motivated though to take some chances and to say "yes" to some doubtful items, the subject's performance could be captured by points B, C, D, E, and F, respectively, depending upon how liberal the subject actually turned out to be when saying "yes." Point F, for example, represents the subject's performance when least cautious; the subject says "yes" a lot and has both a high hit rate and a high false-alarm rate.

Note that any point on a given MOC varies only in the percent correct (or hit rate) and thus the degree of caution used by the subject since the subject's knowledge of the stimuli is fixed at the time of the test. Thus, MOCs are sometimes called isomnemonic functions because a particular MOC represents the locus of all possible points with a single memory strength (Banks, 1970).

Any MOC divides the unit square into two regions: one above it and the other underneath it. Points falling above the MOC represent better performance and those falling below it represent poorer performance. For example, point G in Figure 2 has the same false-alarm rate as point B but a much higher hit rate and hence represents better memory strength than point B. Now since point B is equivalent in memory strength to all other points on the MOC, point G is better than any point on MOC. Similar logic can be used to demonstrate that point H is poorer than any point on the MOC.

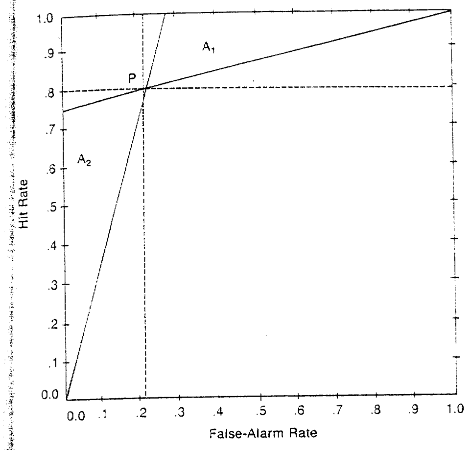
Now the MOC is usually unknown, for its exact shape is a function of the assumed underlying psychological mechanisms and, hence, the as-yet-unknown psychological theory. However, as long as a number of complete, nonintersecting MOCs can be obtained, no theory is needed to determine the relative levels of memory in comparable conditions (Banks, 1970).

The cornerstone in deriving A' as a measure of sensitivity has been the fact that for experiments using the yes-no procedure, the area under the (theoretical) memory-operating-characteristic curve can be interpreted as the percentage correct on all equivalent unbiased forced-choice tests. This holds true for all continuous underlying distributions regardless of their shape (Green, 1964). Pollack and Norman (1964) propose an area measure to convert

(Cont'd on page 31 after Research Currents)

Figure 3

The Representation of a Recognition Performance in the Unit Square



the results of a recognition experiment into an equivalent forced-choice score. More particularly, any point in the unit square can be used to divide the unit square into four regions as shown by the solid lines in Figure 3 where the two solid lines drawn through the target point P, and (0,0) and (1,1) respectively form two nonoverlapping triangles, A_1 and A_2. These triangles define the locus of all possible operating-characteristic curves through the point. Without making any assumptions about the nature of the MOC, Pollack and Norman (1964) chose the ". . . average of the area subtended by the upper, and by the lower, bounds as the measure of recognition performance, i.e., the sum of the I region plus half the A regions."

$$A' = I + \tfrac{1}{2}(A_1 + A_2)$$

In other words, A' is the average of the maximum and minimum possible areas associated with a point.

A computational formula for A' using only a subject's hit and false-alarm rates is given by Grier (1971):

$$A' = \tfrac{1}{2} + \frac{(y - x)(1 + y - x)}{4y(1 - x)}$$

where again x = false-alarm rate and y = hit rate. A' can vary from 0.5 to 1.0 where 0.5 represents chance performance and 1.0 perfect recognition memory.

Using B'_H and A' in Ad Recognition Testing

Note that the parameters B'_H and A' are not directly usable by an individual advertiser interested in assessing the effectiveness of the company's ads. B'_H, for example, measures a subject's response bias over a number of stimuli whereas an advertiser is typically interested in *a* particular stimulus ad. Similarly, A' is not an index of recognition memory for a particular stimulus; rather it measures a subject's ability to discriminate between two classes of stimuli, old (stimulus) versus new (distractor) items. The problem of applying B'_H and A' to individual ads can be resolved though by creating certain new indices.

Consider, for example, B'_H: While B'_H is calculated across all ads presented in a test session, it can be used to determine recognition scores adjusted for response bias for individual ads *by adjusting each ad's score to reflect the response tendencies of those who claimed they had seen it.* To see how this might be done, consider the hypothetical data contained in Table 2.

The example displays the responses for only three ads although there were more ads in the portfolio shown to each of the ten subjects. Each subject's responses to all the ads were used to calculate the person's B'_H parameter. Note that each of the hypothetical ads was claimed to have been seen by one-half of the sample of subjects, although the particular individuals claiming to have seen each ad varied. For convenience purposes, the subjects have been arranged in decreasing order of their "yea saying" tendencies; that is, subject 1 is most inclined to say "yes" and subject 10 is least likely to claim having seen the ad when sorting through the portfolio of actual and distractor ads. When one examines claimed recognition by ad, it seems that the B ad in particular was recognized by those who display a propensity for "yea saying." Conversely, the C ad was recognized by those who do not display this propensity; rather they only respond "yes" when they are fairly sure of themselves. Those recognizing the A ad represent a mixture of response tendencies. *A priori,* it would seem that more subjects "actually saw" the C ad than the B ad even though the claimed recognition for each ad was the same. It would further seem that more subjects saw the A ad than the B ad but that fewer subjects saw the A ad than the C ad.

The indices in the right hand panel of Table 2 capture this sense of the data. Conceptually, the indices are formed by averaging the B'_H values of those who actually claimed recognition of a particular ad. This concept can be easily operationalized by converting the binary responses by subjects into a pair of dummy variables, with yes responses coded 1 and no responses coded 0. The dummy variables are then multiplied by the B'_H values per subject, the products are summed, and the sums are divided by the number of subjects to generate an average adjustment index per ad. Next, response-bias-adjusted scores are computed by adding the product of the average adjustment index and the raw recognition score to the raw recognition score. Note that the adjusted score values are consistent with the *a priori* arguments as to which ads were actually seen the most and the least.

While adjusting raw recognition scores with B'_H is able to correct for response-bias effects, the adjusted scores are still not a true reflection of recognition memory since these scores are also influenced by the nature of the distractor ads used in the recognition test. For example, if the distractor items used in the test were very similar to the stimulus items, it would make the discrimination task harder and would lower the recognition scores. Hence, the raw recognition scores should be adjusted simultaneously by both B'_H and A' indices to account for contamination of recognition scores due to response biases as well as due to the nature of distractors being used.

Such an adjustment can be made by following the same procedure as before. That is, first determine the B'_H values by subject. Then form the three-way products of dummy variable (i.e., 1 if the subject claims to have seen the ad and 0 otherwise) times B'_H value times A' and average these adjusted scores. Finally, add or subtract the adjustment to the raw recognition score for the ad. The index formed thus would logically be called a global-adjusted index because it reflects the true recognition memory of the subject—free of response bias and distractor influences.

The empirical evidence verifying the validity of the proposed indices, but not a discussion of the procedure for making the adjustments, comes from a study by Singh and Churchill (1986) who exposed a group of 80 subjects to a portfolio of 48 printed ads. Subjects were randomly divided into two groups of 40 each. Three weeks later the two groups were shown two different portfolios containing 96 ads each; 48 of these ads were the same as in the original portfolio but the other 48 ads were distractor ads mixed at random. The nature of the distractor ads in the portfolios for the two groups was systematically varied so that in one portfolio the distractor and stimulus ads were quite similar to

Table 2

Unadjusted Recognition Memory Scores and Recognition Bias Adjusted Scores for Three Hypothetical Ads

	Respondent claimed recognition of ad				B'_H for those claiming recognition of ad		
Subject	A	B	C	Subject's B'_H	A	B	C
1	Yes	Yes	No	−0.5	−.5	−.5	0
2	No	Yes	No	−0.4	0	−.4	0
3	Yes	Yes	No	−0.3	−.3	−.3	0
4	No	Yes	No	−0.2	0	−.2	0
5	Yes	Yes	No	−0.1	−.1	−.1	0
6	No	No	Yes	0.1	0	0	0.1
7	Yes	No	Yes	0.2	.2	0	0.2
8	No	No	Yes	0.3	0	0	0.3
9	Yes	No	Yes	0.4	.4	0	0.4
10	No	No	Yes	0.5	0	0	0.5
Unadjusted recognition score	5/10 = .5	5/10 = .5	5/10 = .5				
Sum					−0.3	−1.5	1.5
Average adjusted score					−0.03	−.15	+.15
Response bias-adjusted recognition score					.5+ (.5) (−.03) = .485	.5+ (.5) (−.15) = .425	.5+ (.5) (.15) = .575

one another whereas in the other portfolio they were quite different. Subjects were asked to indicate which ads were stimulus ads and which were not.

Singh and Churchill (1986) argued that since the portfolios seen by the two groups differed only in the type of distractor ads used, the memory for stimulus ads in the two groups should be the same. That is, stimulus ads that are inherently more memorable should be perceived as such by subjects in both groups. Therefore, if the raw recognition scores were adjusted for the response biases of subjects and for the differences in the distractors, then the rank ordering of the stimulus ads with respect to the adjusted scores should be the same in each group. Using B'_H and A' as measures of response bias and memory sensitivity, respectively, they adjusted the raw recognition scores and computed the ranking of the stimulus ads based on unadjusted recognition scores, B'_H adjusted recognition scores, and scores adjusted for both B'_H and A'. The rank order correlations between the unadjusted recognition scores were -0.45, between B'_H adjusted scores were 0.85, and between B'_H and A' adjusted scores were 0.92. In other words, the correlation between the ranking of the ads based on raw recognition scores was extremely poor; it improved dramatically when the influence of subjects' response biases was removed through the B'_H adjustment and improved still further when the differences in recognition due to the differences in distractor ads were removed.

. . . as researchers gather further knowledge of these indices and develop distributional norms across media and product-category type to which to relate the scores, the interpretation problem should be solved.

Thus, it appears that TSD-based measures indeed work. Moreover, they have certain advantages over other proposed methods for determining the recognition of various ads. For example, unlike the methods proposed by Lucas (1942), Simmons (1961), and Appel and Blum (1961), TSD-based measures do not require dual samples and are therefore less expensive to use. In addition, these measures are able to account for distractor differences, which is a basic shortcoming of the methods proposed by Moran (1951a) and Singh and Rothschild (1983). All that is required for their use is the placement of the ads from the test issue in question into a folder with some number of distractor ads. Subjects should be informed about the presence and proportion of the distractor ads in the test. Subjects can then go through the portfolio indicating which ads they have seen before and which ones they have not. From this input each subject's hit and false-alarm rates can be determined as can their B'_H and A' recognition-adjustment indices.

In spite of their merit, these TSD-based adjusted scores are bound to face opposition from some quarters on the grounds that they are not as readily interpretable as are the raw recognition scores. In one way, that position has a great deal of intuitive appeal. An unadjusted recognition score—say that 20 percent of the sample remembered seeing the ad—can be understood by the least sophisticated manager. This is not so with the adjusted scores. Stating, for example, that the adjusted recognition score for the ad was .4 is less compelling intuitively. The index itself is not the problem though; rather it is just researchers' and managers' limited experience in using it.

As experience is gained with these indices, researchers will be able to generate distributions to which obtained values can be referenced. By knowing the frequency with which various values of the indices occur, one will have a measure equally as interpretable as a raw recognition score. For example, the statement that a global index value as high as the one that was observed for a particular ad occurs only rarely, say less than 20 percent of the time, suggests the ad generated a great deal of recognition memory. That is certainly preferable to arguing that 80 percent of the people say they saw the ad using the raw recognition scores, but in reality we do not know how many actually did because of false claiming tendencies. To summarize, as researchers gather further knowledge of these indices and develop distributional norms across media and product-category type to which to relate the scores, the interpretation problem should be solved. The current alternative to not using TSD-based measures is to use invalid recognition measures. ■

References

Advertising Research Foundation. *A Study of Printed Advertising Rating Methods*. New York: Advertising Research Foundation, 1956.

Appel, Valentine, and Milton L. Blum. "Ad Recognition and Respondent Set." *Journal of Advertising Research* 1, 4 (1961): 13–21.

Baddeley, A. D. *The Psychology of Memory*. New York: Basic Books, 1976.

Bagozzi, R. P., and A. J. Silk. "Recall, Recognition, and the Measurement of Memory for Print Advertisements." *Marketing Science* 2, 2 (1983): 95–134.

Banks, W. P. "Signal Detection Theory and Human Memory." *Psychological Bulletin* 74, 2 (1970): 81–99.

Bernbach, H. A. "Decision Processes in Memory." *Psychological Review* 74, 6 (1967): 462–80.

Bower, G. H., and A. L. Glass. "Structural Limits and the Reintegrative Power of Picture Fragments." *Journal of Experimental Psychology: Human Learning and Memory* 2, 4 (1976): 456–66.

Bruce, D., and C. N. Cofer. "An Examination of Recognition and Free Recall as Measures of Acquisition and Long-Term Retention." *Journal of Experimental Psychology* 75, 3 (1967): 283–89.

Campbell, R. A., and L. K. Moulin. "Signal Detection Audiometry: An Exploratory Study." *Journal of Speech and Hearing Research* 11, 2 (1968): 402–10.

Clancy, K. J. "An Exploratory Study of the Validity of the Self-Reporting Technique for Measuring Physician Readership in Medical Publications." New York: *Bio-Medical Publications*, 1974.

———; L. E. Ostlund; and G. A. Wyner. "False Reporting of Magazine Readership." *Journal of Advertising Research* 19, 5 (1979): 23–30.

Dale, H. C. A., and A. D. Baddeley. "Alternatives in Testing Recognition Memory." *Nature* 196, 4849 (1962): 93–94.

Davenport, J. S.; E. B. Parker; and S. A. Smith. "Measuring Readership of Newspaper Advertisements." *Journal of Advertising Research* 2, 6 (1962): 2–9.

Davis, R.; N. S. Sutherland; and B. R. Judd. "Information Content in Recognition and Recall." *Journal of Experimental Psychology* 61, 5 (1961): 422–29.

Deese, J. "Comments on Professor Murdock's Paper." In *Verbal Behavior and Learning: Problems and Processes*, C. N. Cofer and B. S. Musgrave, eds. New York: McGraw-Hill, 1963.

Egan, J. P., and F. R. Clarke. "Psychophysics and Signal Detection." In *Experimental Methods and Instrumentation in Psychology*, J. B. Sidowski, ed. New York: McGraw-Hill, 1956.

Elliott, P. B. "Tables of d'." In *Signal Detection and Recognition by Human Observers*, J. A. Swets, ed. New York: John Wiley & Sons, 1964.

Frazen, R. "Inequalities Which Affect Scores on Advertisements." *Journal of Marketing* 6, 2 (1942): 128–32.

Green, D. M. "General Prediction Relating Yes-No and Forced-Choice Results." *Journal of the Acoustical Society of America* 36, 5 (1964): A1042.

———, and J. A. Swets. *Signal Detection Theory and Psychophysics*. New York: John Wiley, 1966.

Grier, J. B. "Nonparametric Indexes for Sensitivity and Bias: Computing Formulas." *Psychological Bulletin* 75, 6 (1971): 424–29.

Hansen, F. "Hemispherical Lateralization: Implications for Understanding Consumer Behavior." *Journal of Consumer Research* 8, 1 (1981): 23–36.

Haskins, J. B. "Factual Recall as a Measure of Advertising Effectiveness." *Journal of Advertising Research* 4, 1 (1964): 2–8.

Heller, N. "Moran's 'Measuring Exposure to Advertisements.' " *Journal of Applied Psychology* 35, 1 (1951): 77–78.

Hillyard, S. A.; K. C. Squires; J. W. Bauer; and P. H. Lindsay. "Evoked Potential Correlates of Auditory Signal Detection." *Science* 172, 3990 (1971): 1357–60.

Hodos, W. "A Nonparametric Index of Response Bias for Use in Detection and Recognition Learning." *Psychological Bulletin* 74, 5 (1970): 351–354.

Hopkins, R. H., and R. W. Schulz. "Meaningfulness in Paired-Associate Recognition Learning." *Journal of Experimental Psychology* 79, 3 (1969): 533–39.

Klatzky, R. L. *Human Memory: Structures and Processes*. San Francisco: W. H. Freeman & Company, 1980.

Krugman, H. E. "Why Three Exposures May be Enough." *Journal of Advertising Research* 12, 6 (1972): 11–14.

———. "Memory Without Recall, Exposure Without Perception." *Journal of Advertising Research* 17, 4 (1977): 7–12.

Lucas, D. B. "A Controlled Recognition Technique for Measuring Magazine Advertising Audiences." *Journal of Marketing* 6, 4 (1942): 133–36.

———, and S. H. Britt. *Measuring Advertising Effectiveness*. New York: McGraw-Hill, 1963.

Marder, Eric, and M. David. "Recognition of Ad Elements: Recall or Projection?" *Journal of Advertising Research* 1, 6 (1961): 23–25.

Moran, W. T. "Measuring Exposure to Advertisements." *Journal of Applied Psychology* 35, 1 (1951a): 72–77.

———. "A Reply to Heller's Note." *Journal of Applied Psychology* 35, 1 (1951b): 78–79.

Moray, N. "Time Sharing in Auditory Perception: Effect of Stimulus Duration." *Journal of the Acoustical Society of America* 47, 2 (1970): 660–61.

Murdock, B. B., Jr. "An Analysis of the Recognition Process." In *Verbal Behavior and Learning: Problems and Processes,* C. N. Cofer and B. S. Musgrave, eds. New York: McGraw-Hill, 1963.

Neu, D. M. "Measuring Advertisement Recognition." *Journal of Advertising Research* 1, 6 (1961): 17–22.

Parks, T. E. "Signal-Detectability Theory of Recognition-Memory Performance." *Psychological Review* 73, 1 (1966): 44–58.

Pastore, R. E., and C. J. Scheirer. "Signal Detection Theory: Considerations for General Application." *Psychological Bulletin* 81, 12 (1974): 945–58.

Peterson, W. W.; T. G. Birdsall; and W. C. Fox. "The Theory of Signal Detectability." *Transactions IRE Professional Group on Information Theory* P.G.I.T. 2–4, 1954.

Pollack, S., and D. A. Norman. "A Nonparametric Analysis of Recognition Experiments." *Psychonomic Science* 1, 5 (1964): 125–26.

Postman, L. "Choice Behavior and the Process of Recognition." *American Journal of Psychology* 63, 4 (1950): 576–83.

Rilling, M., and C. McDiarmid. "Signal Detection in Fixed-Ratio Schedules." *Science* 148, 3669 (1965): 526–27.

Schmittlein, D. C. "Assessing Validity and Test-Retest Reliability for 'Pick K of N' Data." *Marketing Science* 3, 1 (1984): 23–40.

———, and D. G. Morrison. "Measuring Miscomprehension for Televised Communication Using True-False Questions." *Journal of Consumer Research* 10, 2 (1983): 147–156.

Shepard, R. N., and J. J. Chang. "Forced Choice Tests of Recognition Memory Under Steady State Conditions." *Journal of Verbal Learning and Verbal Behavior* 2, 1 (1963): 93–101.

Simmons, W. R. Abstract of "Controlled Recognition in the Measurement of Advertising Perception." *Public Opinion Quarterly* 25, 3 (1961): 470–71.

Singh, S. N. *Recognition as a Measure of Learning From Television Commericals,* Ph.D. diss, 1982.

———, and M. L. Rothschild. "Recognition as a Measure of Learning from Television Commercials." *Journal of Marketing Research* 20, 3 (1983): 235–48.

———, and G. A. Churchill, Jr. "Using the Theory of Signal Detection to Improve Ad Recognition Testing." *Journal of Marketing Research* 23, 4 (1986): 327–36.

Slamecka, N. J. "Recall and Recognition in List Discrimination Tasks as a Function of the Number of Alternatives." *Journal of Experimental Psychology* 74, 2 (1967): 187–92.

Sorkin, R. D.; R. E. Pastore; and L. D. Pohlmann. "Simultaneous Two-Channel Signal Detection. II. Correlated and Uncorrelated Signals." *Journal of the Acoustical Society of America* 51, 6 (1972): 1960–65.

Starch, D. *Factors in Readership Measurement*. New York: Daniel Starch and Staff, 1946.

Suboski, M. D. "Signal Detection Methods in the Analysis of Classical and Instrumental Discrimination Conditioning Experiments." In *Proceedings of the 75th Annual Convention of American Psychological Association,* 1967.

Sutton, S. "Fact and Artifact in the Psychology of Schizophrenia." In *Psychopathology,* M. Hammer; K. Salzinger; and S. Sutton, eds. New York: John Wiley & Sons, 1972.

Teghtsoonian, M., and R. Teghtsoonian. "Transitory Effect of Number of Alternatives on Performance in a Recognition Task." *Journal of Experimental Psychology* 86, 3 (1970): 467–68.

Tulving, E. "Similarity Relations in Recognition." *Journal of Verbal Learning and Verbal Behavior* 20, 5 (1981): 479–96.

Underwood, B. J. "False Recognition Produced by Implicit Verbal Responses." *Journal of Experimental Psychology* 70, 1 (1965): 122–29.

Van Meter, D., and D. Middleton. "Modern Statistical Approaches to Reception in Communication Theory." In *Transaction IRE Professional Group on Information Theory*, P.G.I.T.–4, 1954.

Weaver, G. E., and C. J. Stanning. "Short-Term Retention of Pictorial Stimuli as Accessed by a Probe Recognition Technique." *Journal of Experimental Psychology: Human Learning and Memory* 4, 1 (1978): 55–65.

Wells, William D. "The Influence of Yea-Saying Response Style." *Journal of Advertising Research* 1, 4 (1961): 1–12.

Zielske, H. A. "Does Day-After Recall Penalize 'Feeling' Ads?" *Journal of Advertising Research* 22, 1 (1982): 19–22.

ERRATA*

Figure 2

The Expanded Parent-Adult-Child Model

Dominator → Biological parents were powerful and supportive.

Authoritarian → Biological parents were remote and critical.

Adult → Biological parents were caring, nurturant, and encouraging of independence.

Adult → Biological parents were caring, nurturant, and encouraging of independence.

Insecure child → Biological parents were powerful and punishing.

Monad → Biological parents were remote and noncommunicative.

* In Figure 2, "The Expanded Parent-Adult-Child Model," appearing in "Parent-Adult-Child Segments in Marketing" by Joseph R. Murphy on page 41 in the April/May 1987 JAR, Vol. 27, No. 2, the middle segments of the figure should have been labeled "Adult," as shown above.