11-1988

Signal Detection Analysis and Advertising Recognition: An Introduction to Measurement and Interpretation Issues

Armen Tashchian
Kennesaw State University, atashchi@kennesaw.edu

J. Dennis White
Florida State University

Pak Sukgoo
University of Tennessee at Martin

ARMEN TASHCHIAN, J. DENNIS WHITE, and SUKGOO PAK*

Recently the theory of signal detection has been introduced as a method for improving ad recognition testing. The authors expand upon this presentation, elaborate the various data collection and analytic approaches available to marketing researchers, and discuss the potential problems associated with each approach.

Signal Detection Analysis and Advertising Recognition: An Introduction to Measurement and Interpretation Issues

Recently, Singh and Churchill (1986, 1987) introduced signal detection theory (SDT) as a method to improve ad recognition testing. Based on statistical decision theory, the application of signal detection analysis to advertising research provides a structured experimental approach that yields a consistent and reliable estimate of the respondent's actual memory for a target ad (i.e., "memory sensitivity"), as well as an estimate of the respondent's decisional bias (in effect, the tendency to over- and underreport recognition of the ad) (Banks 1970; Green and Swets 1966; Swets and Pickett 1982).

Though signal detection analysis is a promising vehicle for such testing, its sizeable empirical literature includes numerous methods for parameter estimation. For instance, typical discussions of the fundamentals of signal detection and the memory assumptions inherent in the theory are based on a data collection method known as the "yes-no task," representative of a class of methods known as "single-interval" paradigms (MacMillan and Kaplan 1985; Singh and Churchill 1987). Researchers in advertising often measure ad recognition through some variant of a forced-choice approach (cf. Singh and Cole 1985; Singh and Rothschild 1983). In the terminology of SDT, these forced-choice methods represent "*m*-interval" paradigms, the most common being the two-interval forced-choice approach (2IFC). These two distinct classes of methods, though both useful for marketing applications of signal detection, involve different data collection demands, require specific interpretations, and lead to different sensitivity estimates. Failure to distinguish between these two paradigms can lead to confusing and inaccurate estimates of advertising effectiveness.

The purpose of our research note is to expand upon previous treatments of SDT in marketing and advertising, elaborate the various data collection and analytic approaches available to marketing researchers, and discuss the potential problems associated with each approach. We first review three specific examples of SDT data collection, drawing a comparison between yes-no tasks and forced-choice methods. Then we discuss the various alternative measures of sensitivity that can result from each approach. Finally, we evaluate the proposed procedures, comparing and contrasting the accuracy of the various measures, and suggest guidelines for their use by ad researchers.

DATA COLLECTION APPROACHES AND ROC ANALYSIS

Typically, SDT data are collected by means of (1) the forced-choice approach, (2) the yes-no task, and (3) the confidence rating method, a variant of the yes-no task. Each involves different experimental procedures and can lead to different measures of memory sensitivity and bias.

Yes-No Task

On each trial in a typical yes-no task, the respondent is presented a single stimulus advertisement (either a real ad, *s*, or a bogus ad, *n*) and is asked to indicate either

*Armen Tashchian is Associate Professor of Marketing, School of Business, Kennesaw College. J. Dennis White is Assistant Professor of Marketing, College of Business, Florida State University. Sukgoo Pak is Assistant Professor of Marketing, College of Business, University of Tennessee at Martin.
The authors thank Charles Hofacker, Judy Vilmain, and two anonymous *JMR* reviewers for their helpful comments.

Journal of Marketing Research
Vol. XXV (November 1988), 397–404

"yes, I remember that ad" (Y) or "no, I do not remember that ad" (N). Instructions typically make explicit mention of the potential presence of bogus ads. The experimenter can select any relative proportion of stimulus and distractor ads, but Ogilvie and Creelman (1968) indicate that equal numbers of stimulus and distractor ads provide the most reliable estimates.

When responding to the stimulus advertisement, respondents can make two types of correct responses: they can correctly recognize an ad to which they previously were exposed (a hit) or fail to recognize a bogus ad to which they were not exposed (a correct rejection). Likewise, two incorrect responses are possible: a reported recognition of a bogus ad (a false alarm) or a failure to recognize a true ad (a miss). Any respondent's performance can be summarized in a 2×2 table where the type of advertisement or stimulus condition is displayed as columns and the recognition response is represented as rows (Figure 1). It is important to realize that the four cells in the matrix can be summarized by two independent values: the hit rate and the false-alarm rate. The hit rate is actually the proportion of real ads correctly identified; thus hit rate = $P(Y|s)$ = hits/(hits + misses). Similarly, the false-alarm rate can be expressed as the proportion of bogus ads incorrectly identified; thus false-alarm rate = $P(Y|n)$ = false alarms/(false alarms + correct rejections).

A critical aspect of signal detection analysis is the realization that hit rates and false-alarm rates will covary when, *ceteris paribus,* decisional bias varies. For instance, one way in which a respondent can report perfect memory is to report recognition of every stimulus ad presented in the test. Though the result is an amazing hit rate of 100%, the respondent's false-alarm rate also is 100%. In turn, the respondent can minimize false alarms by reporting no recognition of any of the ads in the test portfolio (i.e., a false-alarm rate of 0% with a correspondiing hit rate of 0%). The critical contribution of SDT is in acknowledging that the particular decision style adopted by the respondent, either conservative or liberal, is independent of the true level of memory sensitivity and that the two constructs can be captured in separate statistical estimates.

Figure 1
AN ILLUSTRATIVE DECISION MATRIX FOR A YES-NO TASK

		Stimulus Ad	
		Signal (s) (Real Ad)	Noise (n) (Bogus Ad)
Response	Yes (Y)	P (Y|s) Hit Rate	P (Y|n) False Alarm
	No (N)	P (N|s) Miss Rate	P (N|n) Correct Rejection

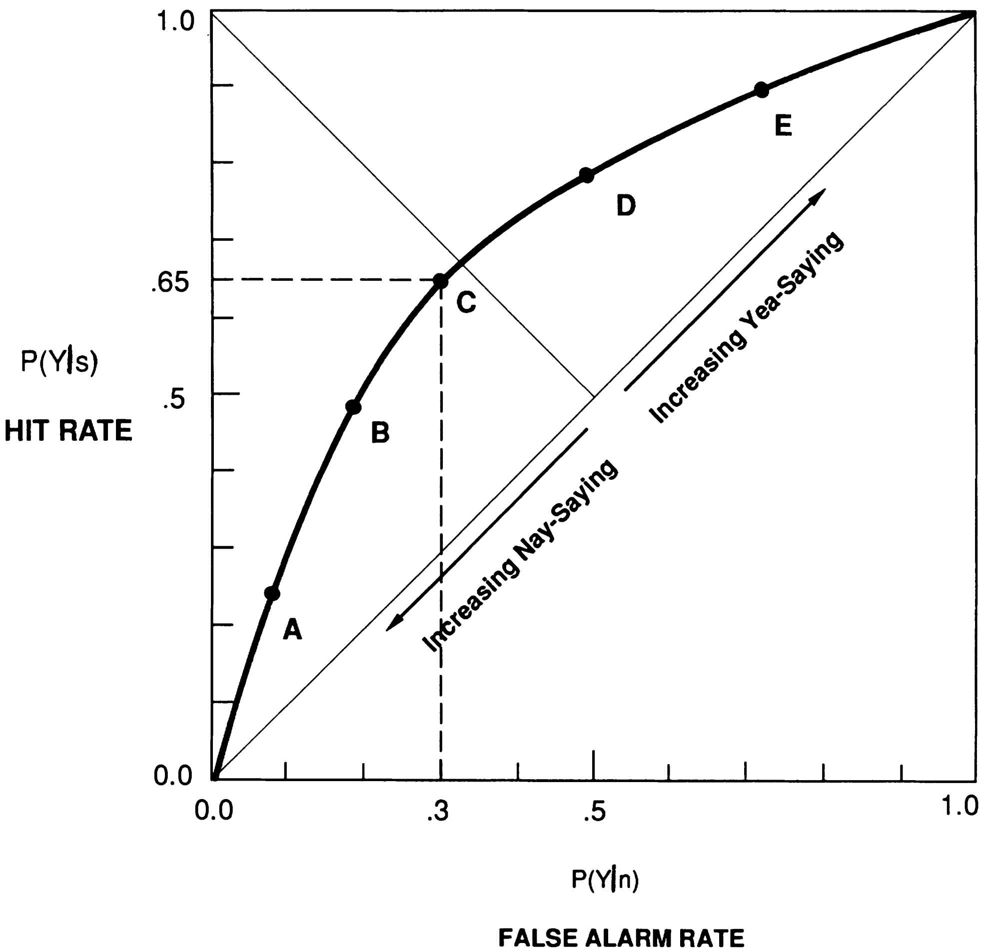
Because the decision matrix can be summarized by two values, it is convenient to represent the respondent's memory-and-decisional performance in a two-dimensional graph generally called the "receiver operating characteristic" (ROC). The ROC is a plot of the hit rate versus the false-alarm rate. Because any respondent's decision matrix represents a single point on an ROC (e.g., in Figure 2, the performance indicated by point C corresponds to $P(Y|s) = .65$ and $P(Y|n) = .30$), the individual must be tested under a variety of decisional criteria to ascertain the form of the entire ROC. One way to do this is to ask the respondent to respond to a block of stimulus items several times, each time becoming either more or less cautious in deciding whether he or she remembers the stimulus advertisement. In effect, the respondent's memory sensitivity remains unchanged while the decision rule used is altered to allow more or less evidence for making a "yes" response. Each new set of trials, and its accompanying decisional criterion, results in a new point on the ROC. Hence respondents can adopt a virtually unlimited number of decisional criteria. In a typical yes-no task, however, respondents are encouraged to adopt three or more such criteria: an extremely conservative response style (in which the respondent must have an unusually high level of certainty before reporting recognition of the stimulus ad), an extremely liberal response style (in which even the slightest feeling of familiarity is sufficient to trigger a recognition response), and one or more intermediate styles.

A few points should be mentioned briefly. First, the respondent's level of decisional bias can be indexed readily on the ROC, either by the normal deviate value of the probability of a false alarm at the criterion (in our case, .3 for point C) or by the slope of the curve at the criterion (Swets and Pickett 1982). These levels generally are denoted Z_k and β, respectively. Second, the positive diagonal represents the locus of points at which false-alarm rates equal hit rates. In other words, performance anywhere along the positive diagonal represents a respondent whose memory is so poor as to be indistinguishable from chance levels of performance.

If we could increase the amount of yea-saying by the respondent, performance would shift toward the upper right corner of the graph. In effect, the closer the performance is to the upper right corner, the greater the "bias toward a positive response" (i.e., yea-saying). Likewise, with increasing bias against a positive response (i.e., nay-saying), performance would shift toward the lower left corner.

The distance of the curve from the positive diagonal indicates the respondent's current level of memory sensitivity. A stronger level of memory sensitivity (regardless of decisional bias) would be represented by any point above the A-E curve. Likewise, poorer memory sensitivity would be indicated by any point below the A-E curve. In summary, the ROC reflects the respondent's memory performance across all possible levels of decisional bias and is therefore independent of that bias.

Figure 2
THE RECEIVER OPERATING CHARACTERISTIC



Confidence Rating Task

To trace accurately the path of the ROC in a yes-no task requires that the respondent rate the stimulus ads several times, each time adopting a different decisional criterion. Because this process can be unrealistically time consuming, psychologists use a "confidence-rating technique" that can generate a complete ROC from respondents within a single test without the necessity of adopting new decisional criteria. Respondents are asked to report their "confidence" in their memory for an advertisement along a k-point scale, where the anchors are "certain I did not see it" to "certain I did see it." These k response categories are used to set up a $k \times 2$ decision matrix, from which $k - 1$ points can be plotted on an ROC.

The procedure is based on the following rationale. We assume responses that are in the highest confidence category result from the respondent applying the strictest possible decisional criterion to the stimulus ads. This case is analogous to a yes-no setting in which the experimenter instructs the respondent to adopt a very conservative decisional criterion: responses to the highest confidence category are counted as "yes" and the remaining categories represent "no." Using the same data, we next assume the respondent adopted a less conservative style; in this case, the respondent would have said "yes" in a yes-no task whenever he or she responded with either of the two highest confidence categories. In effect, the respondent is using a slightly less stringent decisional criterion. We repeat this process cumulatively, across all k categories of the response scale, producing $k - 1$ decision matrices (and hence $k - 1$ ROC points) (see Green and Swets 1966 and McNicol 1972 for details).

As an example, suppose that hypothetical recognition data were collected as shown in Figure 3,A. The cell entries represent the proportions of real ads and bogus ads rated by respondents in each category of the confidence scale. From this initial table, we assume that the respondent mentally establishes four response criteria and,

Figure 3
A HYPOTHETICAL EXAMPLE OF ROC DERIVATION USING A 5-POINT CONFIDENCE RATING TASK

(A)

	STIMULUS AD Real	STIMULUS AD Bogus
5	.27	.05
4	.36	.10
3	.17	.15
2	.15	.30
1	.05	.40

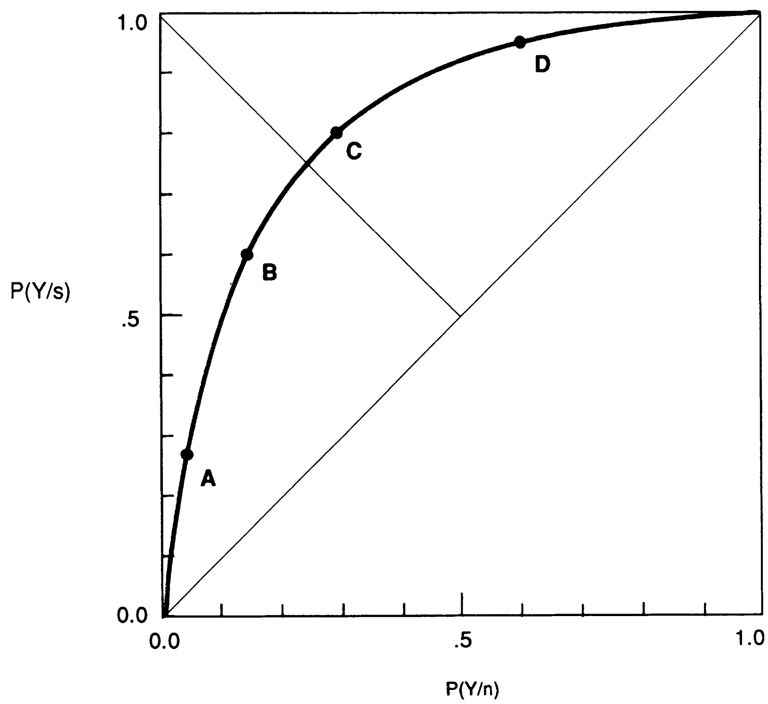
(B)

	STIMULUS AD Real	STIMULUS AD Bogus
C_A	.27	.05
C_B	.63	.15
C_C	.80	.30
C_D	.95	.60
	1.00	1.00

(C)

	C_A Real	C_A Bogus	C_B Real	C_B Bogus	C_C Real	C_C Bogus	C_D Real	C_D Bogus
YES	.27	.05	.63	.15	.80	.30	.95	.60
NO	.73	.95	.37	.85	.20	.70	.05	.40

(D)



(A) Proportion of real and bogus ads in each confidence category
(B)Cumulative recognition proportions listed from highest to lowest level of confidence
(C) The decisional matrices derived from each criterion level
(D)The resulting ROC curve

on that basis, makes recognition judgments as shown in Figure 3,B. Figure 3,C shows the four decision matrices that result. The hit and false-alarm rates derived from these matrices are plotted on the ROC (Figure 3,D).

Memory sensitivity indices. Though an individual's memory performance is represented completely in an ROC, numerical estimates are useful for comparisons across different ROCs. Several alternative indices are available, both parametric and nonparametric.

One specific index of memory sensitivity useful for advertising applications of signal detection is $P(A)$, the area under the ROC. $P(A)$ is a nonparametric statistic that ranges from a lower limit of .5 for chance performance (an ROC along the positive diagonal) to an upper limit of 1.0 for perfect memory performance (Green and Swets 1966). In practice, $P(A)$ is computed by a geometric method for finding the area under the ROC. Generally, with an ROC based on K points,

$$(1) \qquad P(A) = 1/2 \sum_{i=1}^{k+1} [P_i(Y|n) - P_{i-1}(Y|n)][P_i(Y|s) + P_{i-1}(Y|s)]$$

where $P_0(Y|s)$ and $P_0(Y|n)$ represent hit and false-alarm rates of zero and $P_{K+1}(Y|s)$ and $P_{K+1}(Y|n)$ denote hit and false-alarm rates equal to one (McNicol 1972).

An alternative index of memory sensitivity, A_z, is based on past applications suggesting that empirical ROCs are very similar in form to theoretical ROCs that in turn are derived from normal or Gaussian probability distributions (Swets 1986; Swets and Pickett 1982). In practice, if $P(Y|s)$ and $P(Y|n)$ are transformed to z-scores, the normalized ROC will become a linear function. In other words, an individual's memory performance is described adequately by a straight line when plotted on a binormal graph (Dorfman, Beavers, and Saslow 1973; Green and Swets 1966; Swets and Pickett 1982). A_z represents the proportion of the total area of the ROC that is beneath the binormal ROC. Like $P(A)$, A_z ranges from .5 to 1.0.

A final method for deriving a measure of respondent sensitivity involves estimating the area under the ROC with only a single pair of hit and false-alarm rates. Norman (1964) and Pollack, Norman, and Galanter (1964) have demonstrated that a single pair of response rates potentially can provide enough information to determine the approximate path of the entire ROC. Using nonparametric computational formulas (Grier 1971; Pollack and Norman 1964), Singh and Churchill (1986) suggested such a measure of sensitivity can be computed as

$$(2) \qquad A' = 1/2 + (H - FA)(1 + H - FA)/4H(1 - FA),$$

where FA denotes the respondent's false-alarm rate and H is the hit rate. This measure also ranges from .5 (zero or change recognition) to 1.0 (perfect recognition performance) and, like the two aforementioned indices, A' is a pure measure of memory sensitivity, independent of decisional bias.

Forced-Choice Task

The final signal detection method, the forced-choice approach, is a radical experimental departure from the other two methods. In essence, yes-no and confidence rating tasks are what signal detection theorists refer to as "single-interval" paradigms. In other words, the respondent is given a single trial (or interval) containing a stimulus ad that is either signal or noise. In contrast, in forced-choice tasks the respondent simultaneously receives a sequence of stimuli (bogus and real ads) in which only one stimulus is signal and the rest are noise. The respondent is instructed to choose the stimulus most likely to be a signal. A special case of this approach, the two-interval forced-choice (2IFC) procedure, involves the presentation of only two stimuli. Here the real ad is always in either the first or second position (interval) and the respondent is forced to choose one of them.

Moreover, in two-interval paradigms (such as the 2IFC procedure) the respondent is assumed to adopt a decision rule different from what would be appropriate in the single-interval setting. For instance, in the yes-no task we assume the respondent can rank all stimulus ads (real and bogus) along a continuum of subjective familiarity and that the respondent's task is to decide whether the stimulus ad represents a signal or noise. Given underlying normal probability distribution of signal and noise, SDT assumes the respondent determines the ratio of signal to noise distributions where the likelihood ratio is computed as

$$(3) \qquad l(e) = f(e|s)/f(e|n)$$

where $f(e|s)$ is the probability density that the stimulus ad is signal and $f(e|n)$ is the probability density that the stimulus ad is noise. We assume the respondent chooses a particular fixed value of $i(e)$ as a decisional criterion. Any value below this criterion will result in a "no" response and any value equal to or greater than the criterion will lead to a "yes" response.

In contrast, in the 2IFC procedure two events, e_1 and e_2, one corresponding to each stimulus interval, must be compared. The respondent's task is to decide whether the first ad is a signal and the second noise or vice versa. As in single-interval paradigms, the respondent's decision is determined by computation of the likelihood ratio in equation 3. The decision in the two-interval paradigm is complicated, however, by the fact that the respondent must compute two likelihood ratios, one for each interval. The respondent chooses the first interval as signal if and only if the likelihood ratio associated with the first interval is larger than the likelihood ratio associated with the second. Otherwise, the second interval is chosen. The important point is that the two-interval decision rule differs most dramatically from the single-interval rule by not relying on a decisional criterion. The respondent resolves the 2IFC problem by directly comparing two subjective likelihood ratios; no fixed decisional cutoff is necessary. As we discuss shortly, because no decisional criterion is employed in a 2IFC procedure, the use of such a procedure to calculate a measure of respondent bias is questionable.

A sensitivity index for 2IFC. More formally, the two ads presented in each trial in the usual 2IFC method can be viewed as temporally ordered pairs, $<s,n>$ and $<n,s>$, where s denotes a real ad an n a bogus ad. Let $R1$ represent the respondent's decision that the real ad is in the first interval and $R2$ denote the decision that the real ad is in the second interval. $P(R1|<sn>)$ denotes the conditional probability of reporting a signal ad in the first interval when it actually is in the first interval. Likewise, $P(R1|<ns>)$ denotes the conditional probability that the respondent reports a signal in the first interval when the signal actually appears in the second. These two probabilities completely describe the respondent's average behavior in a 2IFC task because

(4) $$P(R1|<sn>) + P(R2|<sn>) = 1$$

(5) $$P(R1|<ns>) + P(R2|<ns>) = 1.$$

The values $P(R1|<sn>)$ and $P(R1|<ns>)$ can be viewed as analogous to hit and false-alarm rates in single-interval paradigms. As Green and Swets (1966) report, these same two probabilities can be used as coordinates of an ROC for the 2IFC task.

A commonly used sensitivity measure for the 2IFC task is $P(C)$ or the proportion of correct decisions. In the most general case, $P(C) = [P(R1|<sn>) + P(R2|<ns>)]/2$. Green (1964) has demonstrated mathematically that $P(A)$, the area under the ROC, is equivalent to $P(C)$ in a 2IFC procedure. Further, Green and Moses (1966) compared both rating-scale measures and 2IFC responses from the same set of respondents. They found that the respondents' forced-choice recognition performance, measured by $P(C)$, was predicted by the area under the ROC calculated by $P(A)$. An excellent discussion of the equivalence of $P(C)$ and the area under the ROC is given by Green and Swets (1966, p. 43–9).

A COMPARATIVE EVALUATION OF SENSITIVITY INDICES

Because each sensitivity measure requires certain data collection methods, problems can arise from combining data collection methods with inappropriate sensitivity indices. In this section, we address three specific questions: Should advertising researchers rely on the A' statistic at the expense of either of the criterion-adjustment measures, $P(A)$ and A_z? Given that A' is the statistic of choice, is it appropriate to compute A' from 2IFC procedures? Can we reasonably use 2IFC procedures to derive a measure of respondent decisional bias?

Comparison of A' with $P(A)$ and A_z

We previously described three general measures of memory sensitivity developed for the single-interval SDT data collection procedures: $P(A)$, A_z, and A'. The three measures bear many similarities. Each represents an estimate of the area under the ROC ranging from chance level of performance (.5) to perfect memory performance (1.0). Each reflects an assessment of respondent memory sensitivity independent of decisional bias.

The three measures differ, however, in terms of data collection demands. The procedures necessary for the standard yes-no task, from which $P(A)$ and A_z would be estimated, involve several replications of the stimulus ad portfolio: one replication for every desired point on the ROC. As it may be unreasonable to expose respondents to multiple replications of various target and bogus ads, use of the A' estimate seems more pragmatic because A' requires only a single presentation of ads. This apparent advantage, however, disappears when we consider the confidence rating technique. By using a 5-point confidence scale, we can derive four points to estimate $P(A)$ or A_z in a single presentation of the portfolio. In terms of data collection, therefore, A' has no particular advantage.

Beyond logistic concerns of convenience, the three estimates differ significantly in their accuracy. Norman (1964) argues that though A' is a more convenient measure, $P(A)$ is clearly more accurate because it is based on several values of hit and false-alarm rates. Moreover, McNicol (1972) reports that computer simulations indicate A' will provide the same values as $P(A)$ only if the respondent is unbiased (i.e., does not exhibit yea-saying or nay-saying). When such biases are present, A' will always result in an underestimation of the true area under the ROC. Because in ad recognition testing we are particularly concerned with the presence of biased responding (yea-saying and nay-saying), A' appears to be the less appropriate measure.

In turn, A_z has certain advantages over $P(A)$. For instance, Swets (1986) argues that the validity of $P(A)$ is highly dependent on the spread of the observed points (hit/false-alarm pairs) along the ROC. Poor placement of the observed points can result in a substantial underestimation of the true area under the ROC (Swets 1986). A_z, in contrast, is calculated by fitting a straight line to the observed data points (plotted on a binormal graph) and is a more efficient, robust measure. In fact, computer programs can efficiently compute the best-fitting line (Dorfman and Alf 1969; Swets and Pickett 1982). For these reasons, A_z is much less dependent on the particular spread of the ROC points. In summary, in both accuracy of estimation and data collection convenience, alternatives to A' are consistently superior.

Computing A' from 2IFC Procedures

Most importantly, each of the three measures traditionally is calculated from single-interval procedures. As discussed before, two-interval methods such as the 2IFC procedure typically lead to the computation of a different statistic: $P(C)$, the proportion correct. This point is critical because the relative ease with which nonparametric measures can be computed, coupled with the attractiveness of forced-choice methods, may lead ad researchers to inappropriate applications of signal detection procedures.

The preceding discussion shows that, for a 2IFC procedure, $P(C)$ is an elegant, efficient way to estimate memory sensitivity and is mathematically equivalent to $P(A)$ as an estimate of memory. A' is, at best, redundant with $P(C)$ as a measure of the area under the ROC. At worst, 2IFC-derived estimates of A' may provide inaccurate measures of ad effectiveness. Given the reservations about sensitivity measures based on a single pair of hit and false-alarm rates (McNicol 1972; Norman 1964), $P(C)$ clearly is the more appropriate measure for a 2IFC procedure.

Computing Decisional Bias from 2IFC Procedures

Researchers may need, in addition to measures of memory sensitivity, an estimate of the respondent's decisional criterion. As discussed before, each respondent in a single-interval procedure is assumed to set a decisional criterion or cutoff in deciding whether the likelihood value should translate into a positive or negative recognition response. In effect, an estimate of this criterion represents the yea-saying and nay-saying tendencies studied by ad researchers. Grier (1971) suggests this measure be estimated as

(6) $B'_H = 1 - FA\,(1 - FA)/H\,(1 - H)$ for nay-sayers

and

(7) $B'_H = H\,(1 - H)/FA\,(1 - FA) - 1$ for yea-sayers,

where H = hit rate and FA = false-alarm rate.

It is important to remember that the respondent does not use a decisional criterion in the 2IFC approach. In fact, 2IFC procedures are used in situations where response bias is an inconsequential issue (Egan 1975, p. 45; Green and Swets 1966, p. 46). To appreciate this point fully we need a clear distinction between *decisional bias* and *interval bias*. Decisional bias, throughout our discussion, refers to a tendency on the part of the respondent to have a consistent preference toward signal or noise. In other words, some respondents are extremely cautious in reporting the recognition of an ad (i.e., nay-sayers) and others are more liberal in reporting recognition of an ad (i.e., yea-saying). This is the very behavior that the decisional criteria in single-interval paradigms are designed to capture. Because respondents in a two-interval procedure do not use a decisional criterion to make a decision, for all practical purposes decisional bias is not present in a 2IFC procedure.

This is not to say that some form of bias does not occur. Two-interval procedures are susceptible to what signal detection theorists refer to as "interval bias." Interval bias reflects a tendency of the respondent to select one interval consistently over the other, a tendency commonly seen in forced-choice experiments. Green and Swets argue that, if strong enough, interval bias can significantly dampen the true magnitude of $P(C)$. However, correction procedures to adjust $P(C)$ for interval bias are simple to compute (Green and Swets 1966).

From a theoretical perspective, the values $P(R1|\langle sn\rangle)$ and $P(R1|\langle ns\rangle)$, though analogous to a hit and a false-alarm rate, are along the negative diagonal of the ROC. The negative diagonal in turn is the locus of points for a respondent having a β equal to one (i.e., the respondent is no more biased to signal than to noise responses). In this case, the use of 2IFC values as input to the B'_H formulas is clearly flawed. In fact, because the respondents' behavior in a 2IFC paradigm is always on the negative diagonal, B'_H values theoretically should equal zero. In practice, computing B'_H from 2IFC results in some value different from zero. However, because B'_H is theoretically equal to zero in a 2IFC setting, any such empirically derived B'_H value can only reflect a test-specific interval bias. B'_H as a measure of decisional bias is appropriate only for data collected by the yes-no procedure.

SUMMARY

Signal detection theory has been applied successfully in several areas in which diagnostic ability must be carefully measured and evaluated, such as medical diagnosis (Swets 1979), military monitoring (Coates, Loeb, and Allulsi 1972), industrial monitoring (Sheehan and Drury 1971), and information retrieval (Swets 1969). It can have equally important applications in advertising research as well as other areas of marketing and consumer behavior (cf. Hutchinson and Zenor 1985; Singh and Churchill 1986, 1987). In marketing, signal detection theory provides a useful and timely paradigm for ad recognition testing. The marketing researcher must understand, however, that signal detection data can be collected under a variety of methodological conditions, can be based on several different implicit models of psychophysical judgment, and can produce several different estimates of memory sensitivity and decisional bias. We present an overview of different data collection approaches in SDT and describe several memory sensitivity indices for each approach. Our review is an attempt to demonstrate the potential confusion and misinterpretation that will result if formulas developed for the single-interval paradigms are transferred to the 2IFC method.

REFERENCES

Banks, William P. (1970), "Signal Detection Theory and Human Memory," *Psychological Bulletin*, 74 (August), 81–9.

Coates, G. D., M. Loeb, and E. A. Alluisi (1972), "Influence of Observing Strategies and Stimulus Variables on Watchkeeping Performances," *Ergonomics*, 15 (July), 379–86.

Dorfman, Donald D. and Edward Alf, Jr. (1969), "Maximum-Likelihood Estimation of Parameters of Signal Detection Theory and Determination of Confidence Intervals-Rating Method Data," *Journal of Mathematical Psychology*, 6 (October), 487–96.

———, Lynn L. Beavers, and Carl C. Saslow (1973), "Estimation of Signal Detection Theory Parameters from Rating-Method Data: A Comparison of the Method of Scoring and Direct Search," *Bulletin of the Psychonomic Society*, 1 (May), 207–8.

Egan, James P. (1975), *Signal Detection Theory and ROC Analysis*. New York: Academic Press, Inc.

Green, David M. (1964), "General Prediction Relating Yes-

No and Forced-Choice Results," *Journal of the Acoustical Society of America*, 36A (May), 1042.
——— and Franklin L. Moses (1966), "On the Equivalence of Two Recognition Measures of Short-Term Memory," *Psychological Bulletin*, 66 (September), 228–34.
——— and John A. Swets (1966), *Signal Detection Theory and Psychophysics*. New York: John Wiley & Sons, Inc.
Grier, J. Brown (1971), "Nonparametric Indexes for Sensitivity and Bias: Computing Formulas," *Psychological Bulletin*, 75 (June), 424–9.
Hutchinson, J. Wesley and Mike Zenor (1985), "Product Familiarity and the Strengths of Brand-Attribute Associations: A Signal Detection Theory Approach," *Advances in Consumer Research*, Vol. 13, Richard J. Lutz, ed. Provo, UT: Association for Consumer Research, 450–3.
MacMillan, Neil A. and Howard L. Kaplan (1985), "Detection Theory Analysis of Group Data: Estimating Sensitivity From Average Hit and False-Alarm Rates," *Psychological Bulletin*, 98 (July), 185–99.
McNicol, D. (1972), *A Primer of Signal Detection Theory*. London: Allen & Unwin.
Norman, Donald A. (1964), "A Comparison of Data Obtained with Different False-Alarm Rates," *Psychological Review*, 71 (May), 243–6.
Ogilvie, John C. and C. Douglas Creelman (1968), "Maximum Likelihood Estimation of Receiver Operating Characteristic Curve Parameters," *Journal of Mathematical Psychology*, 5 (October), 377–91.
Pollack, Irwin and Donald A. Norman (1964), "A Nonparametric Analysis of Recognition Experiments," *Psychonomic Science*, 1 (May), 125–6.
———, ———, and Eugene Galanter (1964), "An Efficient Nonparametric Analysis of Recognition Memory," *Psychonomic Science*, 1 (May), 327–8.
Sheehan, J. J. and C. G. Drury (1971), "The Analysis of Industrial Inspection," *Applied Ergonomics*, 2 (June), 74–8.
Singh, Surendra N. and Gilbert A. Churchill, Jr. (1986), "Using the Theory of Signal Detection to Improve Ad Recognition Testing," *Journal of Marketing Research*, 23 (November), 327–36.
——— and ——— (1987), "Response-Bias-Free Recognition Tests to Measure Advertising Effects," *Journal of Advertising Research*, (June/July), 23–36.
——— and Catherine A. Cole (1985), "Forced-Choice Recognition Tests: A Critical Review," *Journal of Advertising*, 14 (3), 52–8.
——— and Michael L. Rothschild (1983), "Recognition as a Measure of Learning from Television Commercials," *Journal of Marketing Research*, 20 (August), 235–48.
Swets, John A. (1969), "Effectiveness of Information Retrieval Methods," *American Documentation*, 20 (January), 72–89.
——— (1979), "ROC Analysis Applied to the Evaluation of Medical Imaging Techniques," *Investigative Radiology*, 14 (January), 109–21.
——— (1986), "Indices of Discrimination or Diagnostic Accuracy: Their ROCs and Implied Models," *Psychological Bulletin*, 99 (January), 100–17.
——— and Ronald M. Pickett (1982), *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. New York: Academic Press, Inc.

Reprint No. JMR254107

ADVERTISERS' INDEX