SURENDRA N. SINGH, MICHAEL L. ROTHSCHILD, and GILBERT A. CHURCHILL, JR.*

The authors report the second in a series of experiments on recognition as a dependent variable in the study of learning and forgetting of television commercials. They investigate the impact of time since exposure, commercial length, and commercial repetition on recognition and unaided recall scores. The results indicate that recognition scores are not indiscriminately high, as commonly is argued, and that they do decline with time, contrary to what often is assumed. The data, in fact, show that recognition scores are more sensitive and more discriminating than, and covary with, unaided recall scores. The evidence indicates they warrant more consideration by advertisers.

Recognition Versus Recall as Measures of Television Commercial Forgetting

Measuring the effectiveness of advertising is a central research interest of both academic and industry researchers (Leckenby and Plummer 1983; Ostlund 1978; Stewart, Furse, and Kozak 1983). When effectiveness is measured in terms of learning and memory, recall and recognition are common dependent variables (Stewart et al. 1985).

We report the second in a series of experiments on recognition as a dependent variable in the study of the learning and forgetting of television commercials. The first experiment (Singh and Rothschild 1983a) studied *learning curves* as a function of repetition levels, commercial length, and number of distractors; the second experiment examined *forgetting curves* as a function of repetition, message length, and the passage of time. Both studies considered the virtues of recognition as a measure of memory, and its sensitivity and discriminability. The studies also compared recognition with the more popular measure of recall. The two experiments were undertaken in response to the ever-increasing disenchantment of advertising practitioners with recall as a measure of learning (Krugman 1985; Zielske 1982).[1]

The fundamental difference in the two measures is that for *recall* the individual must describe the stimulus, which is not present, whereas for *recognition* the stimulus, which is shown to the subject, must merely be identified as having been seen or heard previously (Bettman 1979). Generally it is easier to show a memory trace through a recognition test than through a recall test. As a result, developing learning to the point where recall can be achieved may be unnecessary if a recognition level of learning is sufficient. The potential implications for media planning and advertising budgeting are important because the cost of achieving an adequate level of learning is much lower for recognition than for recall.

RECOGNITION VERSUS RECALL

Though recognition has been the dominant method used to measure memory of print ads, it has not had the same level of acceptance in broadcast media. Recall measures seem to be popular in broadcast media despite several associated methodological problems (Clancy, Ostlund, and Wyner 1979; Haskins 1964); they tend to be very stringent and can mask the amount of actual memory (Zielske 1982).

*Surendra N. Singh is Joyce C. Hall Faculty Scholar and Associate Professor, University of Kansas. Michael L. Rothschild is Associate Professor and Gilbert A. Churchill, Jr., is Donald C. Slichter Professor in Business Research, University of Wisconsin—Madison.

The investigation was partially supported by University of Kansas General Research Allocation #3026-XO-0038, the Graduate School of Business, University of Wisconsin, and the American Association of Collegiate Schools of Business.

[1]As many of the issues related to the performance of recognition versus recall scores have been discussed at length elsewhere, our treatment is brief. Interested readers are referred to Singh and Rothschild (1983a) for fuller discussion.

Journal of Marketing Research
Vol. XXV (February 1988), 72–80

Whether to choose recognition or recall as a measure of learning of broadcast ads should be dictated in part by the way in which the information is to be used in brand choice decisions. Bettman (1979) believes, for example, that because many brand choice decisions are made at a retail location, particularly for low involvement convenience goods, people only need a recognition level of learning because the buying situation is replete with cues that act as aids to memory. In contrast, if the decision is made at home in isolation, a recall level of learning may be needed because most choices are not physically present in the home for examination. Such would be the situation with high involvement or high brand loyalty products.

Though recognition may be suitable for assessing the memory of low involvement television commercials, its lack of popularity in broadcast media may be due to several causes. First, recognition has been criticized as being less sensitive than recall in that recognition scores are substantially higher than recall scores and can be "indiscriminately high" (Haber 1970; Krugman 1979; Shepard and Chang 1967). Singh and Rothschild (1983a), for example, found recognition scores were significantly higher than recall scores in all cases. In fact, recognition scores exceeded recall scores even when the exposure treatment for recall used more repetitions of the same message or used longer messages. The fact that recognition scores are typically higher than recall scores seems to imply *more sensitivity* because learning exists that could not be shown via recall. Second, advertising recognition scores do not seem to decline with time whereas they should be sensitive to memory loss (Lucas 1960), an issue explored here.

Another criticism of recognition tests is that they cannot discriminate as well as recall tests because of their insensitivity and resulting ceiling effects. However, Singh and Rothschild (1983a) found that when several stimuli were used, the recognition test showed greater variance and fewer extreme scores than did recall.

A final issue related to recall and recognition is whether they measure the same processes of learning. Two general theories are used in psychology to compare recognition and recall. One model, called "strength theory" or "threshold theory," posits that recall and recognition are measuring the same underlying memory construct but that the recognition of an item requires a lower threshold or strength of familiarity than does recall (Kintsch 1970).

A second, more dominant, theoretical model is the "dual-process hypothesis" (Anderson and Bower 1972), which holds that recall involves two steps—memory search and recognition. To recall an item, a subject first generates a number of prospective candidates for recall during the search process and then decides, through recognition, that one of them is the stimulus item. The dual-process hypothesis thus assumes that recognition is a subprocess of recall. Anderson and Bower also have proposed a modified dual-process hypothesis suggesting that recognition and recall involve very similar processes. More specifically, they propose that recognition, too, may involve both search and decision processes (see Mandler 1980 for a detailed review). The empirical evidence supports the notion that recall and recognition scores do covary (Zinkhan, Locander, and Lee 1986).

OBJECTIVES AND RATIONALE

The objectives in the first study (Singh and Rothschild 1983a) were to develop a discriminating recognition measure and to investigate whether recognition could reveal adequately that learning had occurred. These objectives were met in an experiment where the basic content of the advertising stimuli was kept constant and the tests of discrimination were distinguished by the number of repetitions and the lengths of the messages. One result was the demonstration of recognition-based learning curves based on several levels of repetition.

In the study reported here the primary objective was to investigate *forgetting curves* resulting from the passage of time between stimulus exposure and recognition. This issue is an important one, because historically recognition measures have been criticized for their lack of decline across delays in memory tests. In 1956, the Advertising Research Foundation commissioned a study entitled, "A Study of Printed Advertising Rating Methods" (the PARM study). The PARM study found that memory for the ads, as measured by recognition tests, remained stable over a 2-week period. In contrast, recall scores showed significant decline over the same period. This finding was seen as very damaging for recognition testing procedures; researchers such as Lucas (1960) and Wells (1964) argued that because no decline was observed in the recognition scores, recognition was not a measure of memory *per se*. Advertisers became even more reluctant to use recognition measures for assessing learning effects of television commercials when later studies in psychology by Shepard and Chang (1967) and Haber (1970) showed that the human mind could recognize thousands of pictures very accurately after a significant length of time.

Though the belief that recognition scores do not decline over time persists, there are at least two potential explanations for the observed nondecline. One explanation reflects the way recognition tests typically are administered. In advertising, almost all recognition testing is done with yes/no tests whereby the respondent is shown a series of ads one at a time and is asked to respond "yes" if he or she remembers seeing the ad and "no" if he or she does not (Singh and Cole 1985). However, the validity of yes/no recognition tests is questionable. One of the major problems in such tests is false positives that artificially raise the recognition scores. Several studies have shown that a large percentage of respondents claim recognition of bogus ads (ads respondents could not have seen before) when real ads also are being tested (see Singh and Cole for a review of these studies).

The tendency to "recognize" ads irrespective of prior exposure is likely to be due to several factors that affect

ad recognition scores, such as eagerness to please the interviewer, hesitation to appear ignorant, the tendency of people to deny socially undesirable traits and to admit to socially desirable ones, and a generalized trait of some respondents that leads them to overclaim (Clancy, Ostlund, and Wyner 1979; Lucas and Britt 1963). In sum, subjects' response tendencies can affect their claimed recognition of ads.

Another potential explanation for the observed nondecline of recognition scores is theory based. Actually there are two conceptual arguments. One argument holds that recognition memory is generally greater for pictures than for words (Haber 1970; Shepard and Chang 1967; Standing 1973) and is greatest for faces (Deffenbacher, Carr, and Leu 1981). This difference in memory performance may be because the internal representation of a picture contains different kinds of information than does the internal representation of a word (Kosslyn and Pomerantz 1977; Paivio 1971, 1977; Shepard and Podgorny 1978) and faces may be a special class of pictorial material represented in a third memory system (Carey and Diamond 1977).

Another conceptual argument holds that there is a hypothetical memory threshold for recognition that is lower than the hypothetical threshold for recall. Therefore, the strength of a given memory trace may be below the recall threshold, but above the recognition threshold. If the material has been overlearned in relation to the threshold of recognition (or recall), the indicants of recognition (or recall) are not sensitive to early retention losses.

> Thus, when material has been sufficiently overlearned with respect to the recall criterion, recall scores may continue for long periods of time near 100% correct levels. We cannot demonstrate a classical retention curve by the recall measure for the words in the Lord's Prayer or the National Anthem for most adult Americans. A somewhat analogous situation may occur when performance on easy recognition tests shows little or no decline for several days after associations have been learned to the threshold of anticipation. This does not serve as a valid basis for contending, however, that recognition measures, per se, do not yield classical, that is, negatively accelerated, retention curves (Bahrick 1964, p. 190).

The nondecline of recognition scores in the PARM study therefore can be potentially explained in terms of false ad noting by subjects and the use of easy recognition tests that were not sensitive enough to reveal any decline in memory over the short period (2 weeks) after which memory was tested. Recognition memory for advertisements might indeed have a negative slope over time if difficult recognition tests were employed that provide better control of response biases and use a sufficiently longer decay period.

The primary objective of our study was to examine the characteristics of recognition decay as a funtion of time. Two additional objectives were to compare forgetting as measured by the discriminability and similarity of recall and recognition and to compare the data with those of the first study (Singh and Rothschild 1983a) to draw conclusions about the value of recognition as a measurement device for the study of the learning and forgetting of television commercials.

METHOD

To determine whether commercial recognition memory would decline with time if more appropriate conditions were used, we conducted an experiment that reflected the aforementioned three variables. Subjects were exposed to test commercials embedded in a television news show and their recognition memory was tested after different delay periods. Rather than a yes/no test, a forced-choice recognition test was used.[2]

Only two exposures of the test commercials were used to avoid overlearning. Three test periods, 1 week, 3 weeks, and 6 weeks after exposure, were used to make the recognition test discriminating enough for memory loss to be revealed. Memory loss was tested for both 30-second and 10-second spots. The study was a three time periods (1, 3, and 6 weeks) by two lengths of commercial (30- and 10-second) by two repetitions (1 and 2) complete factorial design with multiple recall and recognition dependent variables.

Procedure

Two-hundred five (205) undergraduate student subjects were assigned at random to the 12 experimental cells. The videotapes were shown at three times—at 6, 3, and 1 week before the testing session. At each time, subjects were assigned randomly to one of four 30-minute videotapes. Viewing took place in small groups of 2 to 15 subjects. All post testing was done on a single day to minimize possible contamination due to subject interaction.

After being given instructions that indicated they would later evaluate programming, subjects viewed excerpts of three *network* affiliate news broadcasts. Subjects then were asked to fill out an evaluation form comparing the news from each of the three stations and were asked to come back after 1, 3, or 6 weeks (depending on the experimental condition to which they were assigned) to evaluate *cable* television news shows. This step was part of the cover story to disguise the true purpose of the experiment. At the second session, all subjects were given an unaided recall test of commercials seen previously on

[2]In a forced-choice test, the subject is presented with two or more advertisements at a time. Only one of the ads is the test ad (or stimulus ad); the rest are bogus (or distractor) ads. The subject's task is to pick out the test ad. Because he or she has no choice but to indicate one of the ads is the test ad, the test is called "forced choice." Forced-choice tests are most appropriate in reducing response biases because the subject cannot "please" the interviewer by picking any one of the alternatives. In other words, the tendency to say "yes" affects both the stimulus and distractors alike and hence exerts little influence on the observed choice (Shepard and Chang 1967).

the videotape. Next, subjects received 9-alternative verbal recognition tests of product, brand, and claim recognition.[3]

Stimulus materials: commercials. Three 30-second predominantly informational commercials representing three low involvement categories (trash bags, frozen pie, and salad dressing) were selected. The commericals were edited to create their 10-second versions. The selected commercials and the brands represented in them were novel to the subjects though the commercials were for real brands that were distributed regionally in another part of the United States.

Stimulus materials: overall videotape. Four 30-minute tapes consisting of excerpts of news, weather, and sports from network news shows in different parts of the country were prepared with commercials embedded in the program material in five places. The first two and last two commercials on each tape were nonexperimental fillers used to avoid primacy and recency effects among the experimental commercials. No commercial was shown twice in succession. On two tapes all experimental commercials were 10 seconds long and on the other two tapes all experimental commercials were 30 seconds long. Experimental commercials on tapes 1 and 3 were repeated only once, whereas those on tapes 2 and 4 were repeated twice. Hence, the four tapes captured the two lengths of message (10-second and 30-second) by two repetition levels (1 and 2) aspect of the design. All subjects saw all three experimental commercials.

Dependent Variables

Each subject received an unaided recall test followed by the 9-alternative verbal recognition tests.[4]

Recall test. Subjects were asked to recall the product category, brand name, and claim(s) for as many of the commercials shown to them in the earlier session as possible.[5]

Verbal recognition test. Nine-alternative verbal recognition tests were used for product category, brand name, and claim recognition. Subjects first performed a product category recognition test; next, they were told the actual product category and were asked to pick the stimulus brand name from the nine alternatives. Subjects then were given nine claims and were asked to select the correct claim for each brand. They were not told the correct brand name prior to claim recognition.

RESULTS

Because the study was concerned with recall and recognition as tests of memory rather than the impact of any specific commercial, the results are reported in terms of each subject's aggregate score across commercials. That is, each subject received a score of one for each product, brand, and claim he or she correctly recalled or recognized. Thus, a subject's score for each measure could range from zero to three. An examination of the number of products, brands, and/or claims recalled and recognized indicates that the scores are correlated (see Table 1). This finding suggests that multivariate analysis of variance (MANOVA) is the appropriate analysis technique.

Recognition as a Function of the Passage of Time

One of the strongest criticisms of *recognition* has been its lack of decay over time. Table 2 shows the multivariate and univariate ANOVA results assessing the impact of time since exposure. The results indicate that the passage of time affected recognition scores. The specific impact of time since exposure on recognition can be seen in Table 3. The general pattern is one of decay in recognition scores over time. Though the overall pattern of

[3]Whether recall testing prior to recognition testing affects the observed recognition scores is not known. The psychology literature offers conflicting evidence. Some studies show that prior recall does influence subsequent recognition (Lockhart 1975; Postman, Jenkins, and Postman 1948) and others indicate that prior recall has no effect on subsequent recognition (Dallett, Wilcox, and D'Andrea 1968; Darley and Murdock 1971). We believe that prior recall did not affect recognition scores in this study because Singh and Rothschild (1983b), using forced-choice recognition tests, have shown that prior unaided recall does not influence recognition testing for advertising stimuli.

[4]Recall and recognition are on a continuum of difficulty wherein unaided recall is most difficult and yes/no recognition is most easy in terms of subjects being able to respond to tests of learning. On such a continuum recall can be made easier by adding various cues (aids) and recognition can be made more difficult by adding distractors and by making the distractors similar. By giving strong aids to recall and putting large numbers of similar distractors in a recognition test, one can shift recall and recognition measures along the continuum. In the extreme case, Tulving and Thomson (1973) have shown reversals so that material which is not recognizable can be recalled. The recall test we used is more stringent than that commonly used in industry, where aided recall typically is employed. It seems that the results for recall would have been closer to those for recognition if an aided recall test had been used to measure it.

[5]Though use of aided recall scores is the prevalent industry practice, we chose to use unaided recall measures to keep our dependent variables consistent with those of Singh and Rothschild (1983a), as our study was an extension of theirs.

Table 1
CORRELATION MATRIX OF RECALL AND RECOGNITION SCORES

	Recall			*Recognition*		
	Product	*Brand*	*Claim*	*Product*	*Brand*	*Claim*
Recall						
Product	1.00					
Brand	.44	1.00				
Claim	.72	.37	1.00			
Recognition						
Product	.52	.27	.33	1.00		
Brand	.35	.24	.31	.41	1.00	
Claim	.22	.07	.25	.23	.33	1.00

Table 2
EFFECTS OF TIME DELAY, MESSAGE LENGTH, AND REPETITION ON RECOGNITION AND RECALL

Treatment variable	Criterion variable	Wilks' Λ	Multivariate results F	d.f.	Prob.	Univariate results F	d.f.	Prob.
	Recognition							
Time delay		.882	4.142	6,382	.000			
	Product					9.063	2,193	.000
	Brand					5.073	2,193	.007
	Claim					3.445	2,193	.034
Message length		.903	6.858	3,191	.000			
	Product					.873	1,193	.351
	Brand					5.112	1,193	.025
	Claim					19.264	1,193	.000
Repetition		.822	13.752	3,191	.000			
	Product					35.086	1,193	.000
	Brand					11.513	1,193	.001
	Claim					8.453	1,193	.004
	Recall							
Time delay		.904	3.305	6,382	.003			
	Product					4.268	2,193	.015
	Brand					4.625	2,193	.011
	Claim					6.942	2,193	.001
Message length		.976	1.568	3,191	.198			
	Product					3.976	1,193	.048
	Brand					.118	1,193	.731
	Claim					3.523	1,193	.062
Repetition		.838	12.264	3,191	.000			
	Product					36.819	1,193	.000
	Brand					4.618	1,193	.033
	Claim					19.779	1,193	.000

decay is strong, there are some minor reversals; for example, recognition after a 3-week delay exceeds that after a 1-week delay. In general, though, the frequent criticism that recognition is not susceptible to forgetting is not supported.

Recognition as a Function of Message Length and Repetition

The Singh and Rothschild (1983a) findings indicated that message length and message repetition both can affect recognition scores. Tests of these effects were replicated in the current study; the same results hold in general, though there are a few reversals.

Table 2 shows what happened to recognition as a function of *message length*. Both the MANOVA and ANOVA results show that message length had an impact on recognition with one exception; product recognition did not change when message length changed. This result is somewhat inconsistent with what was found in the previous study, where message length did not affect brand recognition significantly, but did affect product recognition. Table 2 also shows the impact of repetition on recognition. Both MANOVA and ANOVA yield the expected results. In sum, the earlier findings about the impact of message length and message repetition on recognition are supported in the replication, with two exceptions.

Recall as a Function of the Passage of Time, Message Length, and Repetition

The results for *recall* are generally as expected. The MANOVA results for message length are not statistically significant; the ANOVA results indicate that this finding is due to no effect for brand recall. All other relations are statistically significant. The results of the previous study, showing no main effect for length of commercial on brand recall but a significant main effect of length and repetition in all other cases, are replicated perfectly.

Comparing Recall and Recognition

Two major factors to be considered in comparing recall and recognition are (1) which dependent variable is more sensitive to the decay in memory over time and (2) whether the two variables covary.

Sensitivity to memory. Recognition has been maligned for involving substantial ceiling effects and for not being a discriminating measure between stimuli. Shepard and Chang (1967) obtained 85 to 90% recognition scores for stimulus items that had been viewed several weeks prior to testing. However, Singh and Rothschild (1983a) found that the difficult recognition task led to a wide range of scores, a result that was replicated in the current study.

We defined a ceiling effect operationally as being present whenever 15% or more of the subjects could recognize correctly *all three* of the relevant stimuli. Among the 12 cells are six cells in which 15% or more of the subjects could recognize all three *products*. Four of these ceiling effects occurred when subjects were exposed to two repetitions of the ad and were asked for their recognition either 1 or 3 weeks later. There are no ceiling effects for products when recognition was assessed after 6 weeks.

For brand, only one ceiling effect is found among the 12 cells; 15% of the subjects recognized all three *brands* correctly when they saw two repetitions of the ad and their recognition was assessed 1 week later. There are three ceiling effects for *claims*. The total pattern of results seems to suggest that ceiling effects for recognition scores as measured by a 9-item recognition task are not as great a problem for ad testing as commonly has been assumed.

Though recognition scores often have been criticized for ceiling effects, little attention has been directed at examining the behavior of recall scores. In a relatively large number of cells, a large percentage of the subjects could not recall *any* of the three products, brands, or claims to which they had been exposed in the experiment. The situation is just the opposite of that for which recognition measures have been criticized. Consequently a floor effect was defined as present whenever 85% or more of the subjects could not recall *any* of the three products, brands, or claims. The 85% figure was chosen to keep the analysis symmetrical with respect to the operational definition of ceiling effect.[6] Of the 12 cells, a *product* floor effect is present in five, a *brand* floor effect is present in 10, and a *claim* floor effect is present in nine. The conclusion certainly seems warranted that the susceptibility of unaided recall scores to floor effects is greater than the susceptibility of forced-choice recognition scores to ceiling effects.

Another perspective on the sensitivity of the two measures can be obtained by examining Table 3. The data indicate that the mean level of recognition scores for each treatment condition is uniformly higher than that of the recall scores. More dramatically, the mean recognition scores after a 3-week delay exceed the comparable recall scores after a 1-week delay in every case, and the mean recognition scores after a 6-week delay exceed the comparable recall scores after a 1- or 3-week delay in every case. Similarly, mean levels of recognition are higher than mean levels of recall when recognition is assessed after one exposure and recall is measured after two ex-

[6]To be precise, the recognition scores should be corrected for guessing. In a 9-alternative forced choice test this correction is up to 11% (1 of 9), depending on the level of recognition scores. This analysis is not reported because it did not affect a single result in the study.

Table 3

MEANS AND STANDARD DEVIATIONS OF RECALL AND RECOGNITION SCORES FOR THE INDIVIDUAL CELLS

Cell[a]	Message length	Repetitions	Time delay	Recall[b]			Recognition[b]		
				Product	Brand	Claim	Product	Brand	Claim
1 (17)	10 sec.	1	1	.059 (.243)[2]	.059 (.243)	.059 (.243)	1.412 (.795)	1.235 (.752)	1.118 (.600)
2 (16)	10 sec.	1	3	.125 (.342)	.063 (.250)	.000 (.000)	1.125 (1.088)	.750 (.775)	.812 (.655)
3 (20)	10 sec.	1	6	.000 (.000)	.000 (.000)	.000 (.000)	.900 (.912)	.700 (.733)	.850 (.988)
4 (20)	10 sec.	2	1	.800 (.834)	.250 (.444)	.450 (.605)	2.150 (.875)	1.700 (.657)	1.200 (.696)
5 (15)	10 sec.	2	3	.533 (.743)	.000 (.000)	.133 (.352)	1.867 (.990)	1.800 (.676)	1.333 (.488)
6 (16)	10 sec.	2	6	.312 (.602)	.000 (.000)	.062 (.250)	1.312 (.793)	1.125 (.719)	1.250 (.775)
7 (17)	30 sec.	1	1	.294 (.588)	.059 (.243)	.118 (.332)	1.353 (.862)	1.706 (.849)	1.647 (.702)
8 (16)	30 sec.	1	3	.125 (.342)	.000 (.000)	.063 (.250)	1.125 (1.088)	1.250 (.931)	1.375 (.885)
9 (17)	30 sec.	1	6	.118 (.332)	.000 (.000)	.000 (.000)	1.000 (.707)	1.294 (.849)	1.235 (1.033)
10 (20)	30 sec.	2	1	.950 (.887)	.150 (.366)	.650 (.745)	2.200 (.696)	1.600 (.995)	1.900 (.718)
11 (17)	30 sec.	2	3	1.000 (1.000)	.235 (.437)	.412 (.712)	2.353 (.786)	1.529 (.524)	2.000 (.707)
12 (14)	30 sec.	2	6	.429 (.938)	.000 (.000)	.143 (.535)	1.429 (.938)	1.429 (.756)	1.214 (.579)

[a]The number of subjects in the cell is in parentheses.
[b]The standard deviations for the cells are in parentheses.

posures; the results are the same with recognition scores for 10-second commercials versus recall scores for 30-second commercials. Clearly, recognition is more sensitive to learning (as shown in the 1983 study) and to memory rentention over time. Though recall shows weak learning, recognition shows memory retention after six weeks.

Related to the issue of whether there are ceiling or floor effects is the question of the measure's ability to discriminate among subjects in terms of their degree of memory trace of the various commercials. Other things being equal, a measure that shows wide rather than narrow variation across subjects is preferred because it is capable of making finer distinctions. A measure that varies within a narrow band conversely means that subjects having different levels of a memory trace for a commercial would necessarily be judged as having equal levels because of the crude categorization afforded by the scale.

The standard deviation was used to assess the discriminability of the measures; the standard deviations by cell are reported in parentheses in Table 3. Note that for cell 3 the standard deviation of the product recall scores is zero, because none of the subjects could recall seeing any of the three ads. The same was true in six of the cells when subjects were asked to recall brands and in three of the cells when they were asked to recall claims. In contrast, when recognition was assessed, some subjects were always able to recognize at least one of the ads to which they had been exposed.

Table 3 indicates there are very few cells in which the recall scores show more variability than the recognition scores; more specifically, recall scores were more variable in only two of the 12 cells when product memory was being assessed, in no cells when brand memory was being assessed, and in only one cell when claim memory was being measured. If we ignore the cells in which there is no variance in the recall scores, the results suggest that the average ratio of variability of the recognition scores to variability of the recall scores is 1.78 for product, 2.55 for brand, and 1.88 for claim. These ratios understate the true ratios of variation in the two sets of scores because of the necessity of ignoring those cells in which the ratios are infinite. The evidence clearly demonstrates that recognition scores show more variability across subjects than do recall scores, which in turn suggests that they afford finer distinctions among subjects.

Concommitant variation of recall and recognition. The two sets of dependent variables behave similarly in terms of the overall pattern of the results. Spearman's rank order correlation coefficient, for example, indicates that when the 12 cell means are ordered from largest to smallest the rank order correlation between the *product* recall and recognition means is .89 ($p < .05$), whereas that between the *brand* recall and recognition means is .30 (n.s.) and that between the *claim* recall and recognition means is .61 ($p < .05$). Particularly for product, then, but also somewhat for claim, the conclusions would be very similar for the *differential* impact on memory of two repetitions versus one and 30-second versus 10-second exposures regardless of whether subjects' knowledge trace is assessed by using recognition or recall. There would be differences in the conclusions about the *amount* of knowledge subjects had about the products and claims in the commercials when knowledge is assessed by recognition versus recall even though the pattern of effects is similar. One reason for the low correlation between the ranked recall and recognition scores is the low levels of brand recall by cell. The low scores produce only small differences across cells, which in turn generate only a weak ordering of the cells. This finding represents a weak replication of the 1983 finding that all three rank order correlations were significant ($p < .05$).

DISCUSSION

Recognition measures have been used routinely to ascertain the memory effectiveness of print advertisements, but in the past few years researchers have begun to believe that recognition measures should have a greater role in broadcast advertising. For example, many researchers have advocated employing recognition measures for measuring memory effects of television commercials for low involvement products (Bettman 1979; Krugman 1972, 1985; Singh and Rothschild 1983a). Also, dissatisfaction with the currently used yes/no recognition tests seems to be increasing (Singh and Cole 1985). There is a special concern with the response set bias in such tests, which tends to artificially inflate the recognition scores.

Our findings and those of Singh and Rothschild (1983a) show clearly that regulation tests need not have indiscriminately high scores, especially if difficult recognition tests that are less sensitive to subjects' response biases are used and memory is tested after longer intervals. Given that the recognition test used was relatively free of response biases, a nondecline in recognition scores implies that the material being tested has been overlearned for the test being used. One can be reasonably sure that the test is relatively bias free when a 9-alternative forced-choice test is used.

The critical issues for advertising are:

—What amount of learning is adequate? That is, what actual memory threshold level is sufficient?
—Which measure of memory is best under particular circumstances?

In certain low involvement situations where brand choice decisions are made inside the store, a level of learning as indicated by recognition may be sufficient (Bettman 1979). However, even in such situations, many advertisers tend to rely heavily on recall instead of recognition measures.

The issue of which measure of memory is appropriate is important because use of recognition rather than recall may justify lower advertising budgets; a given level of recognition (say 60%) may be reached with far fewer

exposures of an ad than would be necessary if one wanted to achieve the same level of recall. Notice that the *actual* level of memory for a stimulus is constant at a specific time in a particular respondent's mind; it is the response to the measure (recognition or recall) being used to assess memory that varies.

Managers may be overspending substantially as a result of using recall instead of recognition measures. Recall measures seem to understate significantly the true remembrance of advertising in general (Krugman 1972) and of emotional or mood ads in particular (Zielske 1982). Moreover, once the advertising stops, memory as measured by unaided recall declines sharply (Zielski and Henry 1980). In contrast, memory for advertisements declines more slowly when recognition is used to assess it.

The Decay of Memory

Though our findings indicate that recognition memory does decay over time, as had been shown previously in recall tests, strong memory traces remain after long periods of time and after very weak exposure to stimuli. With moderately weak stimulus (two exposures to two 30-second commercials), almost half of the products and brands and more than one third of the claims could be recognized 6 weeks after exposure. In comparison with actual media schedules, this level of exposure can be labeled "moderately weak." Even with the weakest stimulus (one exposure to a 10-second commercial), from one quarter to almost one third of the products, brands, and claims were recognized after 6 weeks.[7]

This demonstration of the power of the human mind to retain relatively inconsequential material is important for advertisers who previously have been shown, via recall, that memory was much weaker. This finding should be useful to members of the *Advertising Age* Sounding Board (Zeltner 1986), who believed that controlling expenditures was the most pressing advertising problem of the late 1980s.

If a recognition level of memory is sufficient to make in-store brand choice decisions (as suggested by Bettman 1979), our findings may be useful to advertisers seeking justification for budget cuts. Memory traces seem to be very long lasting and provide a rationale for more economical flighting schedules.

CONCLUSION

For recognition to be accepted as widely as recall in the assessment of broadcast advertising, it should have certain desirable qualities. It should be sensitive and discriminating, and should measure accurately what is thought to be stored in memory. Our findings and those of Singh and Rothschild (1983a) show that recognition has such qualities.

Recognition seems to be more sensitive than recall. It is able to show memory traces that cannot be seen via recall. Across all cells of the experiment, average recognition is 1.340 items (of three stimulus items) whereas for recall the average is .213 items. In addition, there are 10 cells in which no memory could be measured using recall; the average recognition value is 1.140 items in those cells.

Recognition also seems to be more discriminating. One virtue of a measure is its ability to separate objects on the basis of certain criteria; a key criterion of this ability is the variance obtained across objects on the measure. The variance of recognition is approximately twice that of recall in our study; even this large difference is an understatement, for it does not account for the fact that more than one fourth of the recall cells were not included in the analysis because there was no recall to assess.

Even though recognition is clearly different from recall in some ways, the two measures seem to covary. In the experiment reported here and in the prior one, the rank order correlation of cells across the two measures is high and is statistically significant in five of six cases.

Given two measures of some unknown level of memory, which is more appropriate? A measure can understate reality by being insensitive or can overstate it by allowing false positive responses. Recognition may allow false positive responses, but with a 9-alternative test the upper limit of guessing correctly is 11%. Reducing recognition scores by this amount does not greatly alter the differences between the two measures, but rather strengthens the argument that recognition, if properly used, is a good measure of memory.

Whether one uses recognition or recall as a measure of learning should depend on several factors, including the product type (low vs. high involvement) and place of decision making (in store vs. at home). It also should depend on management's objectives. Managers might prefer a recall level of learning simply because it implies a stronger memory trace for what the advertiser would like consumers to believe about its brand. In this case the information would be readily accessible in a search of memory and an external stimulus would not be necessary to trigger the associations.

Nevertheless, a general conclusion from our study seems warranted: advertisers who in the past have trusted recall as a measure of memory can also have faith in recognition. It is sensitive and discriminating and shows memory loss over time.

[7]Our audience was educated and captive. In addition, the exposure situation involved less clutter than actual TV viewing. Both of these factors might have inflated memory scores over what might typically be found in practice.

REFERENCES

Anderson, John R. and Gordon H. Bower (1972), "Recognition and Retrieval Processes in Free Recall," *Psychological Review*, 79 (March), 97–123.

Bahrick, Harry P. (1964), "Retention Curves: Facts or Artifacts?" *Psychological Bulletin*, 61 (March), 188–94.

Bettman, James R. (1979), *An Information Processing Theory of Consumer Choice*. Reading, MA: Addison Wesley Publishing Company.

Carey, Susan and Rhea Diamond (1977), "From Piecemeal to

Configuration Representation of Faces," *Science*, 195, 312–14.

Clancy, Kevin J., Lyman E. Ostlund, and Gordon A. Wyner (1979), "False Reporting of Magazine Readership," *Journal of Advertising Research*, 19 (October), 23–30.

Dallett, Kent, Sandra G. Wilcox, and Lester D'Andrea (1968), "Picture Memory Experiments," *Journal of Experimental Psychology*, 76 (2), 312–20.

Darley, C. F. and B. B. Murdock, Jr. (1971), "Effects of Prior Free Recall Testing on Final Recall and Recognition," *Journal of Experimental Psychology*, 91, 66–73.

Deffenbacher, Kenneth A., Thomas H. Carr, and John R. Leu (1981), "Memory for Words, Pictures, and Faces: Retroactive Interference, Forgetting and Reminiscence," *Journal of Experimental Psychology*, 7 (July), 299–305.

Haber, Ralph N. (1970), "How We Remember What We See," *Scientific American*, 222 (May), 104–12.

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

Kintsch, Walter (1970), "Models for Free Recall and Recognition," in *Models of Human Memory*, D. A. Norman, ed. New York: Academic Press, Inc.

Kosslyn, S. M. and J. R. Pomerantz (1977), "Imagery, Propositions and the Form of Internal Representation," *Cognitive Psychology*, 9, 52–76.

Krugman, Herbert E. (1972), "Why Three Exposures May Be Enough," *Journal of Advertising Research*, 12 (December), 11–14.

——— (1979), "Low Involvement Theory in the Light of New Brain Research," in *Attitude Research Plays for High Stakes*, J. C. Maloney and B. S. Silverman, eds. Chicago: American Marketing Association.

——— (1985), "Measuring Memory—An Industry Dilemma," *Journal of Advertising Research*, 25 (4), 49–51.

Leckenby, John D. and Joseph T. Plummer (1983), "Advertising Stimulus Measurement and Assessment Research: A Review of Advertising Testing Methods," *Current Issues and Research in Advertising*, 2, 135–65.

Lockhart, Robert S. (1975), "The Facilitation of Recognition by Recall," *Journal of Verbal Learning and Verbal Behavior*, 14, 253–58.

Lucas, Darrel B. (1960), "The ABC's of ARF's PARM," *Journal of Marketing*, 25 (1) 9–20.

——— and Stewart H. Britt (1963), *Measuring Advertising Effectiveness*. New York: McGraw-Hill Book Company.

Mandler, George (1980), "Recognizing: The Judgement of Previous Occurrence," *Psychological Review*, 87 (3), 252–71.

Ostlund, Lyman E. (1978), "Advertising Copy Testing: A Review of Current Practices," *Current Issues and Research in Advertising*, 87–106.

Paivio, Allan (1971), *Imagery and Verbal Processes*. New York: Holt, Rinehart and Winston, Inc.

——— (1977), "Images, Propositions, and Knowledge," in *Images, Perception, and Knowledge*, J. M. Nicholas, ed. Dordrecht, The Netherlands: Reidel.

Postman, Leo, William O. Jenkins, and Dorothy L. Postman (1948), "An Experimental Comparison of Active Recall and Recognition," *American Journal of Psychology*, 61, 511–19.

Shepard, Roger N. and Jih-Jie Chang (1967), "Recognition Memory for Words, Sentences and Pictures," *Journal of Verbal Learning and Verbal Behavior*, 6 (February), 156–63.

——— and P. Podgorny (1978), "Cognitive Processes that Resemble Perceptual Processes," in *Handbook of Learning and Cognitive Processes: V. Human Information Processing*, W. K. Estes, ed. Hillsdale, NJ: Lawrence Earlbaum Associates.

Singh, Surendra N. and Catherine A. Cole (1985), "Forced-Choice Recognition Tests: A Critical Review," *Journal of Advertising*, 14 (3), 52–8.

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

——— and ——— (1983b), "The Effect of Recall on Recognition: An Empirical Investigation of Consecutive Learning Measures," in *Advances in Consumer Research*, Vol. 10, Richard P. Bagozzi and Alice M. Tybout, eds. Ann Arbor, MI: Association for Consumer Research, 271–6.

Standing, Lionel (1973), "Learning 10,000 Pictures," *Quarterly Journal of Experimental Psychology*, 25, 207–22.

Stewart, David W., David H. Furse, and R. P. Kozak (1983), "A Guide to Commercial Copy-Testing Services," *Current Issues and Research in Advertising*, 1, 1–44.

———, Connie Peckmann, Srinivasan Ratneshwar, Jon Stroud, and Beverly Bryant (1985), "Methodological and Theoretical Foundations of Advertising Copytesting: A Review," *Current Issues and Research in Advertising*, 2, 1–74.

Tulving, Endel and Donald M. Thomson (1973), "Encoding Specificity and Retrieval Process in Episodic Memory," *Psychological Review*, 80 (September), 352–73.

Wells, William D. (1964), "Recognition, Recall and Rating Scales," *Journal of Advertising Research*, 4 (3), 2–8.

Zeltner, Herbert (1986), "1986: A Challenging Year for Advertising," *Advertising Age* (January 13), 1+.

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

——— and Walter A. Henry (1980), "Remembering and Forgetting Television Ads," *Journal of Advertising Research*, 20 (April), 7–13.

Zinkhan, George M., William B. Locander, and James H. Lee (1986), "Dimensional Relationship of Aided Recall and Recognition," *Journal of Advertising*, 15 (March), 38–46.