Receiver-operating characteristic curves for taboo and neutral words

JUNE DANDELIKER and DONALD D. DORFMAN, San Diego State College, San Diego, Calif. 92115

Some investigators have applied signal-detection theory to the recognition of such emotional stimuli as taboo words (Dorfman, Grossberg, & Kroeker, 1965; Dorfman, 1967). Although the theory was applied to these stimuli, no good evidence exists to support such an application. The present study evaluated signal-detection theory with respect to its application to the recognition of taboo words. The theory was fitted to rating-method data by means of the method of maximum likelihood, and the chi-square test of goodness of fit was used to evaluate the theory. The data were found to be in reasonably good agreement with signal-detection theory.

Some investigators have recently applied signal-detection theory to the recognition of such emotional stimuli as taboo words (Dorfman, Grossberg, & Kroeker, 1965; Dorfman, 1967). Although the theory was applied to these stimuli, no good evidence exists to support such an application. Application of a theory to a new set of stimuli requires evidence that the theory fits

the data obtained with these new stimuli. The major goal of this study was to determine if signal-detection data obtained with taboo words fit the receiver operating characteristic curves derived from signal-detection theory.

METHOD

The Ss were eight girls enrolled in introductory psychology. Their ages ranged from 18 to 28 years. The stimuli were viewed in a Gerbrands Tachistoscope. For details, see Dorfman, Grossberg, & Kroeker (1965) and Dorfman (1967). Four pairs of words were used: penis, mixer; whore, tally; raped, photo; kotex, cleat. The word-pairs were identical to those used by Dorfman, Grossberg, & Kroeker (1965), and the properties of these word-pairs are described in their article.

The eight girls were randomly divided into four groups of two Ss each. Each group of two Ss was run on one of the four word-pairs. Each S was tested individually in the presence of a female E, and was given 1,500 trials in three sessions of 500 trials. Each session of 500 trials was conducted on a different day, but at the same time of day. On each trial, one of the two stimuli was presented, a taboo word (e.g., penis) or a neutral word (e.g., mixer) for a duration of

50 msec. This duration was found to give an intermediate level of accuracy in the study by Dorfman, Grossberg, & Kroeker (1965). The presentation probability of the taboo and neutral words was equal, and the sequence of word presentations was determined by a table of random numbers. Upon presentation of the stimulus word, the S responded with one of the two words, e.g., penis or mixer, and with a confidence rating (definitely, probably, or possibly). In the instructions, "definitely" was defined as being more likely than "probably," and "probably" was defined as being more likely than "possibly." Before the experiment began, the S was given a card with the two words printed on it. She was told that one of the two words was to be presented on each trial, that she was to respond aloud the word that she thought was presented, and that she was to give a confidence rating of her judgment. Each S was run on one word-pair for the 1,500 trials.

RESULTS AND DISCUSSION

Figure 1 presents the ROC curves for the eight Ss. Ss 1 and 2 were tested on penis-mixer, Ss 3 and 4 on raped-photo, Ss 5 and 6 on whore-tally, and Ss 7 and 8 on kotex-cleat. The solid lines are the theoretical functions fit to the data. These functions were fit by the method of maximum likelihood (see Dorfman & Alf, in press). The fit of the model was evaluated formally by means of the chi-square test of goodness of fit (see Dorfman & Alf, in press). Table 1 presents the parameter estimates for each S and her associated chi square. The σ_n/σ_t is the standard deviation of the neutral word divided by the standard deviation of the taboo word, and $d' = (M_t - M_n)/\sigma_n$. Based upon 95% simultaneous confidence intervals, the results showed that the chi squares of six of the eight Ss were not significantly different from chance. The fit of the model to these data is as good as the fit to the yes-no signal-detection data reported by Swets, Tanner, & Birdsall (1955), and somewhat worse than the fit of the model to their rating-method data (see Dorfman & Alf, 1968; Dorfman & Alf, in press). In the study by Swets, Tanner, & Birdsall (1955), there were four Os, and one of the chi squares was significantly different from chance. This chi square was 7.82, with $df = 3$. They ran about 1,200 trials on each O, so this study was somewhat more sensitive.

Another question of interest is whether the Ss used the confidence ratings in accordance with the theory. If the likelihood ratio of the taboo-probability density to the neutral-probability density is used in setting the criterion, then the

Fig. 1. Receiver operating characteristic curves for the eight Os.

Subject	σ_n/σ_t	d'	Chi Square	df
1	1.08	2.00	9.38	3
2	1.14	1.57	2.32	3
3	0.94	-0.02	2.55	2
4	0.89	1.33	2.16	3
5	1.05	2.34	16.38*	3
6	1.12	1.10	4.01	3
7	1.04	1.61	20.55*	3
8	0.93	0.12	0.40	2

* $p < .05$

probability of the taboo word divided by the probability of the neutral word should increase with the degree of confidence that the word was the taboo word. To determine if the data agreed with this hypothesis, a Spearman rank-order correlation was computed on each S between the above probability ratio and the degree of confidence that the taboo word was presented. In five of the eight Ss, the correlation was unity ($p < .01$), and for the other three Ss, the correlation was positive, but not significant (+.10, +.50, and +.66).

Also of interest is that fact that the sigma ratio (the ratio of the neutral-word standard deviation to the taboo-word standard deviation) averaged 1.02 over the eight Ss. Using the simultaneous 95% confidence intervals for the sigma ratio, based upon the variance estimates taken from the variance-covariance matrix, it was found that, for five of the eight Os, the sigma ratio was not significantly different from 1, whereas for two of the Os, the sigma ratio was significantly greater than 1, and for one of the Os, the sigma ratio was significantly less than 1. These results suggest that the assumption of a sigma ratio of about 1 is a reasonable assumption for these stimuli, at least on the average. For any particular S, the sigma ratio may or may not be 1.

In summary, these data are in reasonably good agreement with signal-detection theory, and suggest that this theory can be applied to answer substantive questions about the recognition of such emotional stimuli as taboo words.

REFERENCES

- DORFMAN, D. D. Recognition of taboo words as a function of a priori probability. *Journal of Personality & Social Psychology*, 1967, 7, 1-10.
- DORFMAN, D. D., & ALF, E. F., JR. Maximum likelihood estimation of parameters of signal detection theory—a direct solution. *Psychometrika*, 1968, 33, 117-124.
- DORFMAN, D. D., & ALF, E. F., JR. Maximum likelihood estimation of parameters of signal detection theory—rating method data. *Journal of Mathematical Psychology*, in press.
- DORFMAN, D. D., GROSSBERG, J., & KROEKER, L. Recognition of taboo stimuli as a function of exposure time. *Journal of Personality & Social Psychology*, 1965, 2, 552-562.

SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. The evidence for a decision-making theory of visual detection. University of Michigan, Electronic Defense Group Technical Report No. 40, 1955.

NOTE

1. This research was supported in part by NSF-GS-1466. These data are part of a Master's thesis by the senior author. Dr. Donald D. Dorfman is now at the University of Iowa.

Meaningfulness ($\bar{m}$): Its relation to frequency and specificity for children aged nine years

NORMA I. MICKELSON, *University of Victoria, Victoria, B.C., Canada*

In this study, Mickelson's (1969) Table of Meaningfulness ($\bar{m}$) for 9-year-old children was examined to determine whether or not meaningfulness ($\bar{m}$) is a function of frequency and specificity. The data indicate that for 9-year-old children, meaningfulness appears to be a function of specificity (χ^2 , 3 df, 11.47, $p < .01$). On the other hand, meaningfulness ($\bar{m}$) does not appear to be a function of frequency (χ^2 , 3 df, 5.6, $p > .05$).

The lack of normative data in the area of children's verbal learning up to the present has prohibited a valid examination of the function of variables that have been shown to facilitate adult verbal learning.

Mickelson (1969) developed a Table of Meaningfulness ($\bar{m}$) for 120 mass and count nouns. Ss were 135 children, aged 9 years, attending school in Saanich, a district adjacent to Victoria, B.C., Canada.

The words for the Table of Meaningfulness were randomly selected from the Thorndike-Lorge list (1944). The sample of words was stratified to control for frequency and specificity. Sixty A words and 60 AA words were included (frequency). In each category, 30 mass and 30 count nouns (specificity) were randomly selected and included. The data revealed that the range of meaningfulness ($\bar{m}$) in the table for 9-year-old children was 8.41-.41, with a median of 4.25 and a mean of 4.29.

The Table of Meaningfulness was then divided into quartiles. On the basis of probability, if meaningfulness is not a function of the variables specificity and frequency, it would be expected that each

quartile would contain an equal distribution of mass and count nouns and of A and AA nouns (Edwards, 1962, p. 63).

Two chi-square analyses of the data were carried out, the first to examine the observed and theoretical frequencies of nouns categorized according to specificity and the second to examine the observed and theoretical frequencies of nouns categorized according to frequency.

RESULTS AND DISCUSSION

The data in this study indicate that those words for which specificity can be defined (count nouns) appeared more often in the high meaningfulness quartiles of the Table of Meaningfulness than would be expected on the basis of probability (χ^2 , 3 df, 11.47; $p < .01$). For children aged 9 years, therefore, it would appear that high-meaningfulness words are more likely to be specific (count) nouns than nonspecific (mass) nouns. On the other hand, the data indicate that the variable meaningfulness ($\bar{m}$) is not a function of frequency (χ^2 , 3 df, 5.6; $p > .05$). One reason for this lack of relationship between high-meaningfulness nouns and frequency, as defined by Thorndike & Lorge (1955), might well be that the Thorndike-Lorge frequency count is not relevant to the language used by today's children.

REFERENCES

- EDWARDS, A. L. *Experimental design in psychological research*. (rev. ed.) New York: Holt, Rinehart & Winston, 1962.
- MICKELSON, N. I. Meaningfulness ($\bar{m}$) indices for 120 mass and count nouns for children aged nine years. *Journal of Verbal Learning & Verbal Behavior*, 1969, 8, 80-82.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.