Criterion shifts and the determination of the memory operating characteristic

K. L. GIBSON
STANFORD UNIVERSITY

Seven university students ran seven sessions in a continuous recognition memory experiment. Responses were made on a 6-point rating scale which was used to construct a memory operating characteristic (MOC). The S's criteria were manipulated by varying the probability distribution of the outcomes, and the effects of the criterion shifts on the MOC were examined. The MOCs for the different conditions were found to be shifted but overlapping.

Egan, Schulman, & Greenberg (1959) attempted to demonstrate that the receiver operating characteristic (ROC) could be obtained economically in the standard signal detection experiment by using a confidence rating scale. If a binary yes-no response is used, the manipulation of the S's criterion is necessary in order to obtain more than one point on the ROC. Kinchla, Townsend, Yellott, & Atkinson (1966) have shown that it is possible for a S to maintain several criteria simultaneously. Therefore, it is possible to obtain more than one point on the ROC in the same session whether a binary response or a rating scale is used.

The use of the ROC has previously been extended to the analysis of the retrieval from memory (Egan, 1958). The present experiment is a combination of the above ideas applied to the memory operating characteristic (MOC). The S's task was to discriminate whether the stimulus presented for testing had been changed from when it was studied or if the stimulus was unmodified. The S made his response on a 6-point rating scale from sure the pair was different to sure the pair was the same. Testing took place in three different locations on the viewer's screen. The three test positions were each associated with a different probability distribution for the two outcomes, stimulus unchanged (same) and stimulus changed (different). The test positions served as cues which could partially determine the criteria in a hypothetical decision process. That is, the experiment attempted to determine how Ss shifted the multiple criteria of a rating scale when the probability distribution of outcomes was changed. The effects of the criterion shifts on the MOC based on the rating scale were examined.

Subjects

Ss were seven Stanford students who were paid $2.00 per session.

Apparatus

Each S was run in a soundproof booth. Stimuli were presented on a cathode ray tube (CRT). Responses were made on a typewriter keyboard beneath the CRT. Generation and presentation of the stimuli and response recording were controlled by an on-line PDP-1 computer.

Stimuli

Each stimulus consisted of two parts. The first part was either a 3, 4, or 5 letter word obtained from Thorndike (1921) with homonyms, personal pronouns, possessive adjectives, and the past tense of verbs eliminated. Each word was used only once in the entire experiment for a given S, and each S received a different presentation order of the stimuli. The second part of the stimulus was one of six two-digit numbers selected for the session. The six two-digit numbers were xy, xz, yx, yz, zx, and zy, with x, y, and z selected at the start of each session without replacement from the numbers 2 through 9.

A stimulus was presented once for study. The S was instructed to try to remember which number was paired with a word. Sometime later in the session the stimulus was tested. The stimulus was either unchanged from when it was studied or it was modified by changing the number to one of the other five possibilities. The S made his response on a 6-point rating scale from sure the stimulus was different to sure the stimulus was the same.

Procedure

Each S was run approximately seven sessions in a two week period. At the beginning of each session eight stimuli as previously described were presented. Then the body of the experiment consisting of alternations between a test phase and a study phase began. A test phase and a study phase together is considered as one trial. There were 204 trials during each session. The experiment is essentially a variation of the continuous procedure reported elsewhere (Atkinson, Brelsford, & Shiffrin, 1967).

Each stimulus was studied and tested once with the major variable being the number of intervening items (study and test phases for other stimuli) between the two events. Since the session began with eight stimulus presentations, there were always eight stimuli which had not been tested. During the test phase a stimulus studied previously in the session but not previously tested was selected at random from all eight such stimuli with the exception that the pair just studied was selected one-half as often as the other possible stimuli. The test consisted of presenting the word TEST and the stimulus. As described before, the stimulus was either the same or different from when studied, and the S made his response on a 6-point rating scale of sameness. The S had up to 5.5 sec to make his response. Feedback was given by flashing SAME or DIFFERENT on the screen for 1.0 sec, which was followed by a blank screen for 1.5 sec. The test

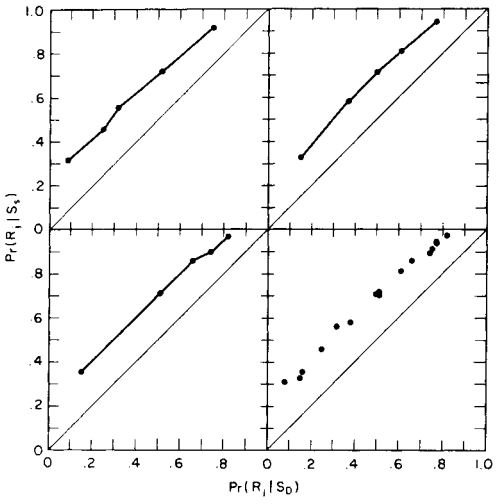

Fig. 1. The MOC constructed from confidence ratings for each probability distribution of outcomes: upper-left, 20% same; upper-right, 50% same; lower-left, 80% same; and lower-right, all conditions.

phase was followed by a study phase during which the word STUDY with a new stimulus (a word never seen before paired with one of the six numbers used for the session) appeared on the CRT for 2.5 sec. The screen was blank for 1.5 sec before the next trial began. In total each trial lasted a maximum of 12 sec.

The test could appear in three possible positions on the CRT. The study phase always appeared in the middle portion. If the test appeared in the upper portion of the screen, the stimulus was the same (unchanged from when studied) 80% of the time; in the middle the stimulus was the same 50% of the time; and in the lower position on the CRT the stimulus was the same 20% of the time. The test position was selected randomly. The S was told that most of the time a pair tested in the upper portion of the screen would be the same, that half of the time it would be the same in the middle, and that most of the time it would be different if tested in the lower portion. The S was instructed to try to use this information in making his response.

Results and Discussion

The data from the first day and the first 30 trials of each subsequent session were deleted from the analysis to allow for warm-up. Lag is the number of trials intervening between the study and test of a given stimulus. Figure 1 presents the MOC constructed from the rating responses with the data from all Ss and lags except zero combined. Each panel represents the data for one test position on the CRT with its associated probability distribution over the outcomes, same or different stimulus. The curves are constructed by considering the possible binary cuts between the rating responses (Green & Swets, 1966, pp. 101-103). The probability of the rating r_j or greater to a same stimulus, $P(R_j | S_S)$, equals $P(r_j | S_S) + P(r_{j+1} | S_S) + \ldots + P(r_6 | S_S)$. Similarly, $P(R_j | S_D) = P(r_j | S_D) + P(r_{j+1} | S_D) + \ldots + P(r_6 | S_D)$.

Although corresponding points shifted over the three test conditions, the curves of the conditions overlapped as shown in the last panel of Fig. 1. The middle confidence ratings indicate that the S was unsure of his response, and it is clear from Fig. 1 that Ss shifted the middle criteria of the rating scale the most over the three test conditions. A large criterion shift in the middle of the scale is for some reason psychologically equivalent to a small change at the extremes of the scale.

Although it is not shown since all lags were combined in Fig. 1, the intercept of the MOC systematically decreased as lag increased. The decreases simply indicated that there was increasing forgetting or interference over lags.

The shape of the MOC obtained from this recognition task differs from that obtained with the more typical old-new recognition task used by Egan (1958). Indeed, from Fig. 1, it appears that the MOC obtained in the present experiment could be well represented by a straight line with slope close to one. The present experiment was an initial attempt to study the feasibility of Ss maintaining several sets of criteria simultaneously. The answer is apparently in the affirmative, but nothing was resolved about the processes involved, and it is not certain that similar results would be obtained with an experimental procedure which resulted in the more typical MOC.

References

ATKINSON, R. C., BRELSFORD, J. W., & SHIFFRIN, R. M. Multiprocess models for memory with applications to a continuous presentation task. *J. math. Psychol.*, 1967, 4, 277-300.

EGAN, J. P. Recognition memory and the operating characteristic. Indiana University, Hearing and Communications Lab., AFCRC-TN-58-51, AD-152650, 1958.

EGAN, J. P., SCHULMAN, A. I., & GREENBERG, G. Z. Operating characteristics determined by binary decisions and by ratings. *J. Acoust. Soc. Amer.*, 1959, 31, 768-773.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

KINCHLA, R. A., TOWNSEND, J., YELLOTT, J. I., Jr., & ATKINSON, R. C. Influence of correlated visual cues on auditory signal detection. *Percept. & Psychophys.*, 1966, 1, 67-73.

THORNDIKE, E. L. *The teacher's word book.* New York: Teachers College, Columbia University, 1921.