THE BISECTION OF A BRIGHTNESS INTERVAL BY PIGEONS¹

R. A. BOAKES

UNIVERSITY OF SUSSEX

Pigeons were trained on a discrete trials, successive discrimination procedure, in which the stimuli were two luminance values on the center key. Behavior was maintained by 25% reinforcement of correct responses on two side-keys. During occasional test trials the luminance of the center key was maintained at one of a number of values, intermediate to those of the two training stimuli, and a function relating the relative frequency of responses on the two side keys to stimulus intensity was obtained. The intersection of this function with the 50% line provided an estimate of the bisection point. Since no bisection point occurred below the geometric mean of the interval, the results were not consistent with a logarithmic scale of brightness but fitted the general mean theorem with an exponent of 0.24. With continued testing, the performance of individual subjects oscillated in an irregular manner about the mean bisection point. The relative stability of the test behavior and the absence of context effects indicated that the method was suitable as a general procedure for measuring stimulus distances.

As a method for scaling psychological distances between stimuli, stimulus generalization studies are in many ways unsatisfactory. The main difficulties are that the effect is sensitive to many non-sensory variables, that it is transient, and that the computation required to convert response rate in extinction into a sensory scale is very complex (Shepard, 1965). In his review of studies in this area, Blough (1966) refers to only two papers that describe alternative approaches to the problem: one is a study by Herrnstein and Van Sommers (1962) and the other an experiment on the bisection of spectral intervals by Honig and Shaw (1962). From a series of experiments (Boakes, 1969) related to the work of Herrnstein and van Sommers it was concluded that their method was not a very promising approach to the study of scaling with animal subjects. However, an incidental product of this research was some data on the bisection of brightness intervals which showed a considerable degree of consistency. This was surprising since Honig and Shaw had found considerable ambiguity in their results and had thrown doubt on the potential of this method.

A relatively complicated way of obtaining bisection points was used by Boakes (1969).

Pigeons were trained to delay a response by one interval in the presence of a bright key and to delay the response by another interval in the presence of a dim key; the stimulus intensity for which either of the delays was equally likely to occur was estimated from an analysis of test trials in which various stimuli of intermediate intensity were presented. It seemed likely that bisection points could be obtained more easily by using simpler response alternatives. In the present study, pigeons were trained to peck one side key if the center key was bright and another side key if it was dim. This behavior was maintained by an intermittent reinforcement schedule and when it had become stable, occasional test trials were introduced. On these trials the illumination of the center key could take one of a number of values and the primary dependent variable was the relative frequency of responses on the two side keys for each test stimulus.

This general procedure resembles the conventional method of transfer testing after training on a successive discrimination. Apart from the fact that this method has been usually applied to studies involving T-mazes or jumping stands, the present procedure differed in two important ways. Firstly, the discrimination was maintained by partial reinforcement and secondly, no reinforcement was available during transfer trials. One purpose of the

¹Reprints may be obtained from the author, Laboratory of Experimental Psychology, University of Sussex, Falmer Brighton, Sussex, England.

present experiment was to discover whether these two factors would lead to more stable test performance than is usual in studies using the conventional procedure.

Since the results of such studies are almost always presented in terms of group means, and the number of test trials for a given stimulus is usually of the order of 10 or 20 (*e.g.*, Sutherland, 1966), there is little information about the behavior of individual subjects. The main reason for this, at least when rats are used as subjects, appears to be that test performance rapidly takes the form of a position habit. Thus, when the results for the first and last halves of a block of tests are compared, the latter is closer to chance performance (*e.g.*, Sutherland and Carr, 1962).

In a few studies, choice data of this kind have been presented for individual subjects (*e.g.*, Scott and Powell, 1963) but no indication is given of any changes that may have taken place during testing.

Bisection studies have not been very popular in human psychophysics since, in common with other forms of category scaling, it is difficult to eliminate from their results the effects of context and stimulus spacing (Stevens and Galanter, 1957). A further purpose of the present experiment was to find out the importance of these effects with the present procedure.

METHOD

Subjects

Eight adult homing pigeons, maintained at 80% of their free-feeding weights, were experimentally naive at the start of the study.

Apparatus

A pigeon chamber of standard dimensions contained three response keys, each 2 cm in diameter. The center key was mounted directly above a grain feeder at a height of 20 cm above the floor. The remaining two keys were mounted on either side of the center key at the same level. The keys were 8 cm apart, center to center. The side keys could be illuminated from behind by a dim red light. The chamber contained no houselight.

A commercial projector was used to illuminate the center key. Between the projector and the key, three neutral density filters and a shutter were mounted independently in a way that would allow them to interrupt the

light beam from the projector when activated by solenoids. The two training stimuli, referred to as Stimulus 1 ("Dim") and Stimulus 8 ("Bright"), were obtained by placing all three filters in the path of the beam (center key luminance, 7.4 ml) or by omitting all filters (center key luminance, 590 ml). The six test stimuli were obtained from the remaining combinations of filters, which gave rise to the following luminance values: 2–15 ml, 3–29 ml, 4–59 ml, 5–74 ml, 6–150 ml, and 7–290 ml. The stimuli were calibrated at regular intervals using an SEI photometer.

Experimental conditions were controlled automatically with standard relay equipment and test stimuli were scheduled by punched tape. For each trial the stimulus, the response, and the occurrence of reinforcement were recorded on a Sodeco print-out counter.

Preliminary Training

The subjects were trained to peck both the white center key and the two red side keys by means of an autoshaping procedure (Brown and Jenkins, 1968). A discrete trial procedure was then introduced, in which a trial started with the center key illuminated. A single peck on this key switched on the red light behind one of the side keys, while the center key remained illuminated, though inoperative. A further single peck on the illuminated side key produced 3-sec access to the grain feeder, which was followed by a 5-sec period of blackout before the next trial. Until a center response and a subsequent side response were made on a given trial, the next trial would not occur. Stimuli 1 and 8 appeared on the center key according to a Gellerman sequence and occurred with equal frequency. The position of the illuminated side key was randomly changed from side to side and was independent of the stimulus present on the center key.

Daily sessions consisted of 100 trials. Four subjects (100% group) received seven sessions under the above conditions. For the remaining four subjects (25% group) the procedure was modified, in that responses on the illuminated side key were randomly reinforced with a probability that was progressively reduced from 1.0 to 0.25 during the first session. Throughout a further six sessions this probability was maintained at 0.25. Side-key responses that were not reinforced were followed by 3-sec blackout preceding the 5-sec intertrial period. Thus, the

time from a side-key response to the beginning of the next trial was always 8 sec, regardless of whether the response was reinforced or not.

Discrimination Training

Discrimination training followed directly after the seven sessions of preliminary training and again consisted of a discrete trial procedure. Each trial started with the center key illuminated with either Stimulus 1 or 8. A single peck on this key switched on the red lights behind both side keys, while the center key remained illuminated. A further single peck on the correct side key produced 3-sec reinforcement and a single peck on the incorrect side key produced a 3-sec period of blackout. As before, trials were separated by a further period of 5-sec blackout and thus a delay of 8 sec occurred between a side-key response and the beginning of the next trial, whether the side response was correct or not. Both a center response and a subsequent side response were necessary in order for the next trial to occur. Before a response on the center key, side-key responses had no effect; after a response on the center key a response on either side key turned off all three lights and made all three keys inoperative until the beginning of the next trial. Daily sessions consisted of 100 trials.

For half the subjects in each group the correct response was to the right in the presence of the Bright stimulus and to the left in the presence of the Dim stimulus ("Bright-right" condition); for the other half these conditions were reversed ("Bright-left" condition). Stimuli 1 (Dim) and 8 (Bright) continued to be scheduled with equal frequency according to a Gellerman sequence.

A total of 10 sessions was given under the above conditions. The average number of trials taken by subjects to reach a criterion of 18 correct out of 20 successive trials was 246, with a range of 131 to 376. The two conditions of preliminary training led to no detectable differences in the rate at which the discrimination was acquired. (A subsequent unpublished experiment using an identical procedure also failed to detect any differences in the acquisition of the discrimination that could be attributed to exposing subjects to intermittent reinforcement during pre-training). On the final session no subject made more than three incorrect responses.

During a further period of six sessions, responses to the correct side key were randomly reinforced with a probability that was progressively decreased from 1.0 to 0.25. The availability of reinforcement on a given trial was independent of which stimulus was present. Correct responses not followed by reinforcement were followed by 3-sec blackout; hence, after the probability of reinforcement had reached its final level, the consequences following 75% of correct responses were identical to those following incorrect responses. During the same period the number of trials in each session was increased to 200. Little change occurred in the percentage of incorrect responses either during this transition or in a final five sessions, where the probability of reinforcement following a correct response was maintained at 0.25, that were given before the test procedure began. Thus, just before testing, each subject was receiving approximately 50 reinforcements per session.

Testing Procedure

At this stage of the experiment the subjects were again assigned to two groups of four: Group A contained "Bright-right" subjects, P79 and P81, and "Bright-left" subjects, P93 and P95. Group B contained "Bright-right" subjects, P119 and P121, and "Bright-left" subjects, P120 and P122. In each of the above pairs the preliminary training of one subject had been under the 100% condition and that of the other under the 25% condition. This assignment was made as a precaution against what seemed a remote possibility that differences in the preliminary training might still show some effect.

Conditions during testing were identical to those of the preceding stage except that test trials were introduced, on average, every fifth trial. The number of normal trials between successive test trials ranged from four to eight and no test trials occurred within the first 20 trials of each session. For Group A, any of the test stimuli could be present on a test trial and their order was randomized within the constraints that each stimulus was presented six times per session and occurred only once within a block of six successive test trials. For Group B, only three test stimuli were used: 4, 5, and 6. Each was presented 12 times per session and only twice within a block of six successive test trials. Test trials were scheduled

independently of the stimuli during normal trials and also independently of reinforcement.

The only difference between a test trial and a normal trial, apart from the change in the luminance of the center key, was that no reinforcement ever occurred on a test trial. Group A was given 24 test sessions and Group B 30 test sessions.

During both training and testing procedures sessions occurred daily, except for a break of a few days after the tenth session of discrimination training.

RESULTS

Psychometric Functions

Psychometric functions were obtained for each subject by plotting the percentage of left responses to each stimulus as a function of stimulus intensity. These are shown in Fig. 1, where the percentages are calculated over the first three test sessions. Thus, for Group A each point is based on a total of 18 presentations of a particular test stimulus and for

Group B on a total of 36 presentations. The functions in this figure suggest a continuous, monotonic relationship between the frequency of responding to a given side and stimulus intensity. When a logarithmic scale of stimulus intensity is used, as in Fig. 1, this relationship is approximately of a normal ogive form. This emerges more clearly in the averaged functions shown in Fig. 4. The greater regularity in the functions for Group B presumably reflects the fact that the percentages for this group are based on twice as many trials per test stimulus as those for Group A. These functions are based on only three successive test sessions because changes, which are discussed below, occurred as testing continued.

Effects of Prolonged Testing

On the basis of previous studies employing transfer tests, it was expected that the frequency of a given response to a test stimulus would drift steadily towards either 0 or 100% as testing was continued. In fact, only two subjects (P119 and P122) exhibited a steady drift of this kind, though it is possible that if testing had been continued very much longer all subjects would eventually have shown a strong preference for one side or the other on test trials. However, within the present testing period, which exceeded 850 test trials for each subject, the most common form of performance was an irregular oscillation. This is shown in Fig. 2 where the percentages of left responses to Stimuli 4 and 6 are plotted for successive blocks of three sessions. For the sake of clarity the corresponding points for the other test stimuli are omitted, but these followed the same general pattern. The frequencies of left and right responses to the two training stimuli remained unchanged throughout the test period and the average frequency of incorrect responses on normal trials did not exceed 2% for any subject.

Since responses were never reinforced on test trials, whereas correct responses on normal trials continued to be reinforced 25% of the time, the test procedure is a form of discrimination training. The effect of this was an increase in latency, i.e., the interval between the beginning of a trial and the first response on the center key, on test trials. This is also shown in Fig. 2, where the difference between average latencies on normal trials and average latencies on test trials is plotted for each subject. In

Fig. 1. Percentage of responses to the left key over the first three test sessions as a function of stimulus intensity. The percentages are obtained from a total of 18 trials per test stimulus in the case of Group A and 36 trials per test stimulus in the case of Group B.

GROUP A

GROUP B

NUMBER OF TEST SESSIONS

Fig. 2. Percentage of left responses to Stimuli 4 and 6 on successive blocks of three test sessions and differences between average latencies on normal and test trials as a function of number of test sessions. The percentage of left responses to Stimulus 4 (filled circles) and to Stimulus 6 (open circles) are based on 18 trials for Group A and 36 trials for Group B. The latency differences (crosses, joined by broken lines) represent the average difference between the mean latency to all test stimuli and the mean latency to training stimuli over three successive sessions.

comparing these changes in latency for Group A and Group B, the difference in absolute values is not very informative, since the test stimuli for Group B are on average a greater distance from the training stimuli than those for Group A. What is important is that the latency differences for Group B reach a maximum after 12 to 18 sessions and then decrease. This means that no practical limit on the number of test sessions with this procedure is likely to be imposed by the extinction of responding during test trials. Presumably, responding on test trials was maintained by the requirement that both a center-key and a side-key response were necessary for the next normal trial to occur. Furthermore, in Group A the latency differences for only two birds (P79 and P81) appear to have reached a maximum during the test period of 24 sessions. Thus, increasing the number of different test stimuli appears to reduce the rate at which the discrimination between test and normal trials develops. The average interresponse times between the first response on the center key and the first response on a side key were also recorded and compared on normal trials and on test trials. Average interresponse times on test trials increased slightly with three subjects as testing continued, but the differences were about one tenth of the latency differences.

From the relative frequencies to each stimulus over successive blocks of three test sessions the bisection point was estimated for each subject (see below) and the median values for each group are shown in Fig. 3. Despite large changes in the performance of individual subjects, shown in Fig. 2, these median functions are remarkably stable for the first 15 sessions; beyond this point it was no longer possible to estimate bisection points for two subjects in Group B, and thus no medians are shown for this group. The apparent shift of the medians for both groups after the first three sessions from the brighter end of the range towards the geometric mean was not found to be significant ($p < 0.10$). This was tested using a Wilcoxon signed rank test to compare the bisection point for the first three sessions with the median value for the following three blocks of three sessions.

Effects of Context

Two types of context effect were investigated. The first was a possible sequential ef-

Fig. 3. Average position of bisection points for Groups A and B for blocks of three successive sessions as a function of number of test sessions.

fect, by which performance on a test trial might depend on the stimulus present on the preceding normal trial. For example, if the Bright stimulus had been presented on the previous trial, a response appropriate to the Dim stimulus might be more likely to occur to the next test stimulus. A separate analysis was made of test trials that followed the Bright stimulus and test trials that followed the Dim stimulus over the first three test sessions. This is shown in Fig. 4 where the average relative frequencies of "Bright" responses (right responses for "Bright-right" birds and left responses for "Bright-left" birds) are plotted as a function of stimulus intensity for each group. A sequential effect of the kind described above would displace to the right the function for test trials following normal trials with the Bright stimulus. As can be seen, there is no sign of any displacement in either direction.

The second context effect concerns the spacing of the test stimuli. When human

Fig. 4. Performance of Groups A and B on the first three test sessions represented as the frequency of "Bright" responses for each test stimulus when test trials followed normal trials in which (a) the Bright training stimulus and (b) the Dim training stimulus were presented.

observers are used as subjects in category scaling tasks, there is a marked tendency for test stimuli to be equally distributed among the available categories (Stevens and Galanter, 1957). Since the test stimuli for Group B are on average brighter than those for Group A, the presence of any bias due to stimulus spacing should shift the bisection point for Group B towards the brighter end of the range. A slight difference in this direction was detected, as is seen in Fig. 4, but this was not found to be significant.

Estimation of Bisection Points

A common question that arises in studies of this kind concerns the optimal number of test trials for estimating the pertinent psychophysical parameter. If very few trials are considered, estimates are likely to be inaccurate, whereas the use of a large number of trials will include the variability introduced by the effects of the testing procedure. The parameter of interest in the present study is the bisection point of the range defined by the training stimuli, *i.e.*, the luminance value of the center key for which subjects would make an equal number of left- and right-key responses. This value was obtained by plotting psychometric functions for each subject of the kind shown in Fig. 1, using a normal probability scale for the ordinate. When only the percentage of left responses to Stimuli 3, 4, 5, and 6 were used for Group A and to all three test stimuli

Fig. 5. Changes in the standard deviation of the bisection point positions and also changes in the difference between the average positions for Groups A and B, when the estimates were based on an increasing number of test sessions.

(Stimuli 4, 5, and 6) for Group B the functions became more linear than those shown in Fig. 1. Straight lines were fitted by eye to these points and the point of intersection of each linear function with the 50% line was obtained. This was done for the first session, the first two sessions, and so on; the variance of these values between subjects was found to decrease to a minimum when the first three sessions were considered and then to increase as further test sessions were included. This is shown in Fig. 5, where the standard deviation over all eight subjects is plotted as a function of cumulated sessions. Also shown in this figure are the differences between the mean bisection points for Groups A and B. In view of the variance changes shown in Fig. 5, bisection points were estimated from the first three test sessions and the values for each subject are shown in the left-hand column of Table 1.

With respect to human psychophysical studies, Fagot (1963) has shown that if one assumes a power function relationship, of the form $P = kI^n$, between perceived magnitude (P) and physical intensity (I), the value of the exponent (n) in this function can be obtained from bisection data using the general mean equation:

$$2 I_{1/2} = I_B^n + I_D^n$$

In the application of this formula to the present results, I_B and I_D represent the intensities of the Bright and Dim training stimuli and $I_{1/2}$ the intensity of the obtained bisection point. If the bisection point were at the *arithmetic* mean of the two training intensities the value of n would be 1.0; if they were at the *geometric* mean, n would be equal to zero. The values of the bisection points shown in the left-hand column of Table 1 were substituted into the above equation and the corresponding values of the exponent, n , were obtained using an iterative solution. These values are shown in the right-hand column of Table 1.

Table 1

Points of bisection of the range defined by the Dim (7.4 ml) and the Bright (590 ml) training stimuli and the corresponding exponent from the general mean theorem.

Subject	Bisection point in ml	Exponent
P79	120	0.26
P81	66	0
P93	123	0.27
P95	71	0.03
P119	138	0.33
P121	110	0.22
P120	123	0.27
P122	112	0.23

DISCUSSION

The problem of determining bisection points is a particular application of the procedure employed in this study. As with traditional transfer tests, the set of stimuli employed could just as well be complex as from a single physical dimension. For example, the general properties of the procedure were not altered when a set of 12 patterns was used (in an unpublished experiment by Boakes and Sutherland). The relative stability of the group averages, when compared to either generalization tests in extinction or transfer tests, allows the possibility of using this as a baseline to assess the effects of changes introduced during the test period. For example, the effect of stimulus spacing discussed earlier might be investigated by changing the set of stimuli after the start of the test period, thus allowing within-group comparisons to be made. Alternatively, if a complex set of stimuli were used, one could investigate the effect on

test performance of altering some attribute of the training stimuli.

The rate of reinforcement for normal trials employed in this study was chosen for convenience; there was no indication during the training phase that reducing the probability of reinforcement after a correct response below 25% would present serious problems and it seems likely that at least a further small reduction would increase the stability of the test behavior.

With regard to the particular problem of bisection, an important aspect of the results is the absence of sequential effects and the small effect of stimulus spacing, since both present major problems in comparable human studies. The possible reasons for this include the use of lengthy training and testing, instead of verbal instructions and relatively short testing periods, and the method of interspersing a number of normal trials between successive test trials instead of massing the tests; also, human observers may be just more complicated psychophysical subjects. Of these, the spacing of test stimuli seems likely to be the most important factor and it is proposed to check this guess by testing human subjects with a similar procedure.

Using a conditioned suppression technique, Shumake, Hatfield, and Smith (1966) showed that brightness thresholds for White Carneaux pigeons are proportional to intensity for the range of intensities used in the present experiment and thus obey Weber's law. In their study, subjects were required to discriminate between the presence and absence of a white circle on a white background of slightly different intensity. This differs from the present procedure, but it seems highly probable that Weber's law would also hold for threshold measurements using a discrimination between intensities that changed from trial to trial as in the present experiment. This implies that a scale of perceived brightness based on discriminability units or *jnd's* should be logarithmic function of stimulus intensity, as in Fechner's law. Such a scale would predict that the bisection point should occur at the geometric mean of an interval. The present results and Boakes' (1969) indicate that pigeons tend to bisect brightness intervals on the brighter side of the geometric mean. Hence, the Fechnerian logarithmic function appears to provide an inappropriate scale for bisection data.

An alternative scale is provided by a power function. The median exponent for such a function, as determined by the general mean equation, was 0.24 for the present subjects. This is somewhat higher than the value of 0.15 obtained in Boakes (1969). These earlier results were more variable than the present ones and this may account for the difference. An alternative explanation is provided by a recent study of bisection using human subjects (Stewart, Fagot, and Eskildsen, 1967) where it was found that exponents were reduced when either the size of the bisected interval was increased or the absolute value of the brighter end of the interval was increased. The interval used in the present experiment (7–590 ml) was both smaller and at a lower intensity level than the interval (60–65 $\times 10^3$ ml) used in Boakes (1969). If this latter account is correct it means that either a power function is also inconsistent with bisection data or that no uniform scale can be obtained by this method. A study of the effects of varying bisected intervals over a large range of intensities is needed before accepting bisection as a satisfactory method for direct scaling with animals.

REFERENCES

- Blough, D. S. The study of animal sensory processes by operant methods. In W. K. Honig (Ed.), *Operant behavior: areas of research and application*. New York: Appleton-Century-Crofts, 1966. Pp. 345-379.
- Boakes, R. A. Response continuity and timing behaviour. In R. M. Gilbert and N. S. Sutherland (Eds.), *Animal Discrimination Learning*. London: Academic Press, 1969. Pp. 351-384.
- Fagot, R. F. On the psychophysical law and estimation procedures in psychophysical scaling. *Psychometrika*, 1963, 28, 145-160.
- Herrnstein, R. J. and Van Sommers, P. Method for sensory scaling with animals. *Science*, 1962, 135, 40-41.
- Honig, W. K. and Shaw, J. The bisection of spectral intervals by pigeons: a first attempt. Paper read at Eastern Psychological Assoc., April 1962.
- Scott, T. R. and Powell, D. A. Measurement of a visual motion after-effect in the rhesus monkey. *Science*, 1963, 140, 57-59.
- Shepard, R. N. Approximation to uniform gradients of generalization by monotone transformations of scale. In D. Mostofsky (Ed.), *Stimulus generalization*. Stanford: Stanford University Press, 1965. Pp. 94-110.
- Shumake, S. A., Hatfield, C. A., and Smith, J. C. Brightness difference thresholds in the pigeon using the conditioned suppression technique. *Psychonomic Science*, 1966, 6, 313-314.
- Stevens, S. S. and Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, 54, 377-411.
- Stewart, M. R., Fagot, R. F., and Eskildsen, P. R. Invariance tests for bisection and fractionation scaling. *Perception and Psychophysics*, 1967, 2, 323-327.
- Sutherland, N. S. Partial reinforcement and breadth of learning. *Quarterly Journal of Experimental Psychology*, 1966, 18, 289-301.
- Sutherland, N. S. and Carr, A. E. Visual discrimination of open and closed shapes by rats. II. Transfer tests. *Quarterly Journal of Experimental Psychology*, 1962, 14, 140-156.

Received 1 July 1968.