How repeatable are Stevens's power law exponents for individual subjects?*

MARTHA TEGHTSOONIAN and ROBERT TEGHTSOONIAN†
Smith College, Northampton, Massachusetts 01060

Magnitude estimations of apparent length and apparent area were obtained for the same group of Ss over successive experimental sessions. Session-to-session correlations between individual exponents on a given continuum were positive and reliable for successive 24-h intersession intervals, but were not significant for a 1-year interval. In a second experiment on judgments of apparent area, when each stimulus was judged only once per session, the session-to-session correlation was reliable only when the intersession interval was zero. Six other intervals, ranging from 1 to 77 days, yielded nonsignificant correlations. When the constraints exerted by repeated judging are removed, the location of S's exponent in a distribution of exponents is stable only for brief intervals. Thus the differences among exponents cannot reflect any persisting attributes of Ss' sensory or judgmental processes.

There is now much evidence that the psychophysical power law, $\psi = k\phi^n$, describes the data of individual Ss as well as group averages; Marks and Stevens (1966) have summarized data supporting this conclusion for a wide variety of perceptual continua. But the existence of variability in individual exponents poses a problem of interpretation: Is the variability due to random error or does it reflect characteristic individual differences in sensory or judgmental processes? If the variability is random, each S's exponent can be regarded as an estimate of the "true" exponent for the continuum in question and can, therefore, be pooled with exponents obtained from other Ss to improve the estimate. But if the variability reflects genuine individual differences, then an exponent derived from pooled data may serve only as a convenient summary statistic, an average of all the individual scores.

The question can be approached by measuring exponents for a group of Ss on two successive occasions. If the variability is indeed random, session-to-session correlations between individual exponents should be zero. However, Jones and Woskow (1962) and Ekman et al (1968) have reported low, but significant, correlations between exponents for two stimulus continua when the same response continuum is used, and Ekman et al have reported high correlations when there is no change in either stimulus or response continuum.¹ However, we have no information on the effect of number of successive sessions or interval between sessions, both variables which might be expected to influence the probability that S

repeats a previously made judgment and, thus, to affect the correlation between exponents. If these correlations reflect some basic sensory or judgmental characteristic of individual Ss, they ought to be relatively insensitive to practice effects or variation in intersession interval.

Two experiments were designed to investigate these variables. In the first, Ss made magnitude estimations on 5 successive days of both the apparent length of straight lines and the apparent area of polygons. As a first step in determining the effect of intersession interval, Ss were retested 1 year later. In the second experiment, Ss judged the apparent area of circles in two sessions, but the interval between sessions varied from a few minutes to 77 days.

EXPERIMENT 1

Method

Subjects. Ss were 14 undergraduate women.

Materials: Lengths. An aluminum beam, 1 x 1/4 in. and 72 in. long, was mounted horizontally at a height of 44 in., 6 ft away from S. Two black markers slid back and forth to demarcate line segments of specified length, which were always symmetrical around the center of the beam. Five different sets of 9 lengths were used. Each set was selected from a total set of 26 lengths, ranging from .50 to 60 in., so that it would contain as few values as possible in common with the other four sets.

Materials: Areas. Irregular polygons of black matte cardboard were presented against a white felt background, 34 in. square, 4 ft from S. There were five different sets of 9 areas. Each set was selected from a total set of 26 areas—each represented by a different polygon—ranging from 2.5 to 600 sq in., so that it would contain as few values as possible in common with the other four.

Procedure. Each S served for five sessions, scheduled on 5 successive days. On each occasion, S judged both lengths and areas. The five sets of lengths and of areas were assigned randomly to sessions, except that each S judged all sets. At the first session, S was instructed to assign numbers according to the apparent lengths or areas (see M. Teghtsoonian, 1965, for a discussion of instructions for apparent size as opposed to objective length and area). No standard was designated or modulus assigned. The nine values of a particular set were presented twice in random order. The same procedure, with abbreviated instructions, was followed at subsequent sessions.

One year later, 10 of the original 14 Ss who were available for retesting judged again the same sets they had first judged.

Results and Discussion

Quality of the data. Confirming the results of Stevens and Guirao (1963) and M. Teghtsoonian (1965), the data, when pooled over Ss and sessions, are well fitted by power functions, and the exponents of .93 for apparent length and .77 for apparent area are typical of previous findings. Furthermore, power functions are excellent descriptions of the data for individual Ss. The straight line fitted by least squares to log judgment vs log length accounts for an average 98% of the variance in an individual S's judgments, while for area the comparable value is 97%.

Stability over sessions. Analysis of variance showed no effect of session on exponent for either apparent length or area ($p > .20$ for both), although the effect of Ss was significant ($p < .01$ for both). Thus, while Ss differ reliably in their exponents, the experience of repeated judging exerts no effect on the mean value of these exponents.

Another possible effect of repeated judging is that S may become better at the task. For apparent area, the portion of variance in log individual judgments accounted for by the fitted straight lines increased from 94% in Session 1 to 98% in Session 5, a significant improvement ($p < .001$, according to an analysis of variance). For apparent length, the comparable values varied between 98% and 99% for all sessions, but the changes were not significant ($p > .20$); perhaps this was a case where there was no room for improvement.

In neither analysis was the effect of Ss significant ($p > .10$ in both). The goodness of the power functions as a fit to individual data appears to be a random variable.

Session-to-session correlations. Figure 1 shows the Pearson

*This research was supported by USPHS Grant MH 11810-01.

†We thank Jane Newman and Marilyn Ingellis, who ran Ss for Experiment 1, and Tobey Levine, who prepared stimulus materials and assisted in data analysis for Experiment 2.

Fig. 1. Correlations r between individual power function exponents obtained in successive sessions. The bottom abscissa shows the session—from 1 to 6—and goes with the points connected by dashed lines; these points represent correlations between exponents on different continua, length vs area. The top abscissa shows pairs of sessions—1 and 2, 2 and 3, and so on, separated by 1 day, and Sessions 1 and 6, separated by 1 year—and goes with the points connected by unbroken lines; these points represent correlations between exponents on the same stimulus continuum, area in one case and length in the other. For Sessions 1-5, which were separated by 1 day, $N = 14$; for Session 6, which was 1 year later, $N = 10$.

product-moment correlation coefficients r between individual exponents for Session n and Session $n + 1$ for apparent length and apparent area. All are moderately high and positive (a mean value of $+0.84$ for apparent length and $+0.76$ for apparent area) and significant ($p < .05$). At any stage from the second to the fifth successive session, S 's exponent is closely related to his performance on the previous day. In conjunction with the significant effect of S s on exponents, this suggests that there are indeed characteristic values for S s that endure from one day to the next. The effect of judging repeatedly may extend over long intervals, since the correlation between Session 5 and Session 6 1 year later is $+0.63$ ($p < .05$) for apparent area, although only $+0.27$ ($p > .05$) for apparent length. However, a better estimate of the relatedness of exponents following a long interval is the correlation between Session 1, the first session, and Session 6, 1 year later, since it is less influenced by the repeated judging in Sessions 2-5. The values are shown in Fig. 1; for apparent area, r is $+0.40$ and for apparent length, $+0.36$ ($p > .05$ in both cases).

The correlations between exponents for apparent length and apparent area for the same session are also shown in Fig. 1. For Sessions 1 and 5, the correlations were not significant ($p > .05$), while for Sessions 2-4, they were moderately high and positive (mean value $+0.70$) and significant (in each case, $p < .05$). The correlation for Session 6, 1 year later, was even lower than in Session 5 and was not significant ($p > .05$). Thus, S 's judgments of the two continua are unrelated in his first session and again on the last of five successive sessions, as they remain 1 year later. The relatedness exhibited in Sessions 2-4 is evidently a transitory phenomenon.

In summary, Experiment 1 found session-to-session correlations between individual exponents for the same stimulus continuum but also showed that their magnitude may depend on the length of the intersession interval; Experiment 2 was designed to explore this variable. There is also evidence that S may be constrained to use similar judgments despite differences in the class of stimuli presented, but such constraints were transitory in nature in this situation. They were minimal or absent when S made his first judgments and when he had made frequent repeated judgments, and they may also have depended on the length of the intersession interval.

EXPERIMENT 2

Experiment 2 was designed to determine the size of intersession interval over which session-to-session correlations between exponents for the same stimulus continuum are reliable. Groups of S s judged the apparent area of circles in two separate sessions, with the interval between sessions ranging from 0 to 77 days.

Method

Subjects. S s were students in an introductory laboratory course in psychology; seven sections of 13 students each constituted the seven groups of the experiment. Since not all S s appeared for Session 2, final N s were 12, 10, 11, 10, 11, 9, and 10 for groups with increasing delays.

Materials. Outline circles of nine areas, $.028$ – 28.3 sq in., were reproduced by Ditto and assembled into booklets in 26 random orders. Each booklet contained initial pages of instructions.

Procedure. All groups received Session 1 at the beginning of a class period during the same 1-week interval. One group received Session 2 at the same time. Session 2 was given to the other groups, at the beginning or end of a class period, at the following intervals after Session 1: 1 day, 2 days, 1, 2, 8, and 11 weeks.

Fig. 2. Correlations r between individual power function exponents when the interval between two sessions varied. N ranged from 9 to 12 at different intervals. Only the correlation for no delay is significantly different from zero.

The instructions preceding the first set of nine circles asked S to assign numbers proportional to the apparent areas of the circles; to discourage S from judging diameter or calculating area, different irregular polygons were shown as examples of variations in apparent area. Neither standard nor modulus was provided. The instructions preceding the second set of nine circles were an abbreviated version; they neither stressed, nor cautioned against, consistency with the first set of judgments.

Each S had a different random order of areas for Session 1 and Session 2. Within a group, each S had different random orders from all other S s; each group had the same set of random orders as all the other groups.

Results and Discussion

Quality of the data. Power functions were fitted to the data of each S . The mean exponent, averaging over groups and sessions, was 0.69, a value in good agreement with previous results. The mean value of r^2 for individually fitted functions was 0.96, again confirming the excellent fit provided for the judgments of individual S s.

Session-to-session correlations. Pearson product-moment correlation coefficients r were calculated between the individual exponents obtained in Session 1 and Session 2 for each group and are shown in Fig. 2. Only in the case of the zero-delay group is r ($+0.82$) significant ($p < .05$).

Thus, the session-to-session correlation is a remarkably short-lived phenomenon, not even extending over the 24-h interval for which it was observed in Experiment 1. The fact that all but one of the obtained r s were positive, although not reliably different from zero after a 1-day interval, indicates that some of the variance may be associated with individual S s over the period up to 11

weeks. But the average value of $r = +.3$ indicates that, at most, a rather small portion of the variance in individual exponents is thus accounted for.

This experiment does not tell anything about how the correlations in exponents observed in Experiment 1 and for the zero-delay group in Experiment 2 come about. But we may speculate that repeated judging has at least two effects: (1) S remembers specific stimuli and repeats remembered judgments in the interest of consistency, and (2) even in the absence of explicit recognition of stimuli, S tends to repeat previously made responses. Conditions which favor retention—e.g., repetition of judgments, short time intervals, distinctive stimuli and experimental setting—will also favor high correlations.

This may explain why session-to-session correlations were more evanescent here than in Experiment 1 (where a 1-day interval yielded a moderate correlation for apparent area exponents). In that experiment, S judged two sets of stimuli, each set twice in a single session; in this experiment, S judged one set of stimuli only once in a session.

DISCUSSION AND CONCLUSIONS

Previous investigators of session-to-session correlations between individual exponents have argued that these imply enduring characteristics (sensory or judgmental) of individual Ss. But our data, while confirming the existence of such correlations, reveal their dependence on short intersession intervals. The 1-year interval of Experiment 1 yielded correlations of $+.3$ and $+.4$ for apparent length and area exponents, respectively; in Experiment 2,

there is an average correlation of $+.3$ between apparent area exponents for intervals of 1 day to 11 weeks; none of these values reached statistical significance. Taken together, they suggest that at least 90% of the variance in individual exponents is attributable to chance factors when the constraints introduced by repeated judging are minimized. This contrasts with the estimate offered by Ekman et al (1968) of about 10%, a value which reflects their procedure of running both sessions in a single day.

Results from Experiment 2 support the conclusion that one set of judgments can constrain a subsequent set when the interval between them is very short. And it may be that, if several judgments of each intensity are made in a single session, the extent of constraint is extended to as much as 24 h. In summary, there appear to be real individual differences in exponents, but the factors that produce them are stable for such short periods of time that they cannot be regarded as enduring individual characteristics. These results are for apparent size, and how far they may be generalized we do not know.

These findings have practical implications for procedure in number-matching experiments. In general, if E wants independent estimates of exponents, he should let at least 1 day elapse between sessions and have S judge a stimulus set only once per session. The practice of some Es of having S judge the same set of stimuli repeatedly or of having S make judgments of two or more continua in a single session permits previous judgments to constrain subsequent ones, producing individual exponents which are not independent across sessions or between continua.

Furthermore, the high correlations between sets of judgments close together in time mean that the amount of information in second and succeeding sets is small. Increasing the number of stimulus values presented and decreasing the number of repetitions would yield more information for the same investment of time.

Finally, there is evidence to support the practice of using highly practiced Ss. Not only may their judgments be less erratic (relative to fitted power functions), but their exponents for different continua judged in the same session show greater independence than is found for less experienced Ss.

REFERENCES

- EKMAN, G., HOSMAN, B., LINDMAN, R., LJUNGBERG, L., & ÅKESSON, C. A. Inter-individual differences in scaling performance. *Perceptual & Motor Skills*, 1968, 26, 815-823.
- JONES, F. N., & WOSKOW, M. On the relationship between estimates of magnitude of loudness and pitch. *American Journal of Psychology*, 1962, 75, 669-671.
- MARKS, L. E., & STEVENS, J. C. Individual brightness functions. *Perception & Psychophysics*, 1966, 1, 17-24.
- STEVENS, S. S., & GUIRAO, M. Subjective scaling of length and area and the matching of length to loudness and brightness. *Journal of Experimental Psychology*, 1963, 66, 177-186.
- TEGHTSOONIAN, M. The judgment of size. *American Journal of Psychology*, 1965, 78, 392-402.

NOTE

1. Ekman et al correlated, not exponents, but scores—the ratio of S's judgments for the strongest and weakest stimuli—which they regard as approximations of exponents for individual Ss.

(Accepted for publication December 26, 1970.)