A note on the up-and-down method

CHARLES HANLEY[1]
MICHIGAN STATE UNIVERSITY

Objections to the use of the original version of the up-and-down method in psychology, as exaggerating individual differences and yielding discrepant results, overlook statistical, psychometric, and logical factors. When the method is applied incorrectly, the standard error of SD is large and apparent differences in estimates of SD are not statistically significant. Psychometrically, a single-trial procedure should give less reliable, hence more variable, PSEs than one using multiple trials. But this effect increases standard error of the mean, thus significant mean differences found with the method are not attributable to greater variability in PSEs. Explanations of discrepancies in results from up-and-down and traditional methods in the study of perceptual development fail to account for numerous similarities. Finally, a novel method ought to produce some novelty, if it is to make a contribution to psychology.

The original form of the up-and-down method (Dixon & Mood, 1948), which allows estimation of mean, SD, and their confidence limits on the basis of one observation per S, appears especially suited to the study of perceptual development, since the single-trial procedure eliminates any role for age-related response biases that arise in methods using repeated trials (Hanley & Zerbolio, 1964). The single trial, moreover, comes closer than the stately methods of traditional psychophysics to approximating the situation holding for real-life judgments of variables like length and size. Giving up multiple measurements on each S, of course, loses precision that can only be recovered by employing large numbers of Ss, but the up-and-down method concentrates judgments close to the mean, thereby requiring fewer Ss than would be the case with single-trial versions of the methods of limits or constant stimuli. In escaping known response biases, however, the original up-and-down procedure also surrenders the possibility of obtaining PSEs for individuals and patently is useless in studies where such data are wanted.

Recently, Wohlwill (1968) has criticized the use of the original version of the method in psychological research, partly for doing what best recommends it: "the method fails to reveal the overconstancy bias found in adults by other methods . . . ," "there is no opportunity for him (the S) to become effectively adapted to the situation, to establish a frame of reference for his judgments, etc.," and "(the method) yields perceptual judgments such as would be obtained in the absence of exposure or adaptation to the situation." Moreover, other passages imply that a single-trial method somehow restricts the amount of time S has for his judgment, when in fact the only restriction is that there be just one judgment.

But these considerations perhaps are mainly matters of taste for research problems. Of the two fundamental objections Wohlwill raises, one concerns the effect of a single-trial procedure on the SD of individual PSEs, a statistic the method does calculate. The psychological factors involved in a single-trial procedure, according to Wohlwill, are responsible for "highly variable judgments." His supporting data come from a combined up-and-down method and method of adjustment study of the bisection of distance. On the up-and-down trial, SD of PSEs was 15.28 cm, while on two following adjustment trials, SD was 4.29 cm. The difficulty with his up-and-down result, however, is that his estimate of SD is highly undependable. The method requires that the interval (d) between levels of the variable stimulus be neither too coarse nor too fine. When the interval is too coarse, the estimate of the mean suffers; when too fine, SD becomes unreliable. The method yields useful data when the ratio of d to the SD found by the method falls between 0.5 and 2.0. In Wohlwill's study, which corresponds to a pilot attempt to locate a suitable value of d, the d/SD ratio

was 0.16. To use this estimate of SD as evidence that a single-trial procedure exaggerates individual differences, presumably beyond utility, ignores the question of confidence limits for the estimate.

As d/SD falls below 0.5, estimates of SD by the up-and-down method become increasingly unreliable, as shown by their standard errors. The graph needed for computing the standard error of SD does not go below a d/SD ratio of 0.20. With that value instead of the actual 0.16, and the assumption that the method otherwise was optimally applied, the standard error for the obtained SD of 15.28 cm is 6.13 cm, and 5% limits for SD are 3.27 and 27.29 cm. SD obtained from adjustment trials falls within these limits, so that statistically there is no significant difference between the two estimates of SD. Rather than demonstrating that a single-trial method produces a "highly inflated picture of the inter-S variability," Wohlwill's study shows that deviating from the recommendations regarding d/SD yields an undependable estimate of SD.

On psychometric grounds, PSEs based on a single trial should be less reliable than PSEs based on the average of several trials. For this reason, the SD of single-trial PSEs should be somewhat larger than the SD of average PSEs, given that true PSEs do not change from trial to trial. In choosing d, in the absence of a pilot study with the method itself, an investigator would be wise to aim for a d/SD ratio of 1.0 to make allowances for the reliability factor, rather than risk a ratio approaching 0.5.

The second fundamental objection Wohlwill raises has to do with the comparability of results from single-trial and multiple-trial methods. Part of the objection relates to his attempt to explain the discrepancy between estimates of SD in the bisection study by appealing to psychological processes like adaptation and practice. Given that line of reasoning, which overlooks statistical and psychometric interpretations, it is natural to attribute any discrepancy involving estimates of *means* to the same set of psychological factors. The empirical data he cites come from a developmental study of five illusions by Hanley and Zerbolio (1965) where, according to Wohlwill, there is a "failure to replicate the marked reduction of the Müller-Lyer illusion with age which has almost invariably been found in previous studies . . ."

This objection needs further examination. First, a large SD does not bias an estimate of the mean, or make it unreliable. The standard error of the mean is the relevant reliability statistic. If SD is overestimated, standard error of the mean will be too large, and significant differences will fail to appear when real differences exist. But when significant differences do appear, they cannot be attributed to psychological factors responsible for an exaggerated SD. Second, Wohlwill overlooks instances where the Hanley and Zerbolio study agrees with other investigations. With the Müller-Lyer illusion, for example, their data indicate that the illusion declines with age after six years, just as in most other reports. The discrepancy Wohlwill cites holds for Ss under six, who are not so often studied, yet the closest comparable information, Noelting's (1960) age trends on a first Muller-Lyer adjustment trial, also shows less susceptibility in younger as compared to slightly older Ss. Furthermore, up-and-down method results for the Ponzo illusion fit precisely with other reports (Leibowitz & Heisel, 1958; Leibowitz & Judisch, 1967) in showing a steady increase in susceptibility to the illusion from age four to adulthood.

Logically, an explanation of discrepancies between single-trial and multiple-trial procedures should also allow for cases of agreement between the two kinds of methods. This elaboration is lacking in Wohlwill's analysis. And, finally, to condemn a method on the grounds that it can produce a somewhat different picture of perceptual development than is customary with methods Wohlwill elsewhere severely criticizes (1960, p. 251) is not likely

to help understanding of processes of perception. Economy aside, what reason is there to use a novel method, if not to disclose what traditional approaches may sometimes conceal?

REFERENCES

DIXON, W. J., & MOOD, A. M. A method for obtaining and analyzing sensitivity data. *J. Amer. Statist. Assoc.*, 1948, 43, 109-126.

HANLEY, C., & ZERBOLIO, D. J. Application of the up-and-down method in measuring perceptual development. *Child Develpm.*, 1964, 35, 979-984.

HANLEY, C., & ZERBOLIO, D. J. Developmental changes in five illusions measured by the up-and-down method. *Child Develpm.*, 1965, 36, 437-452.

LEIBOWITZ, H., & HEISEL, M. A. L'évolution de l'illusion de Ponzo en fonction de l'âge. *Arch. Psychol.*, 1958, 36, 328-331.

LEIBOWITZ, H., & JUDISCH, J. M. The relation between age and the magnitude of the Ponzo illusion. *Amer. J. Psychol.*, 1967, 80, 105-109.

NOELTING, G. La structuration progressive de la figure de Müller-Lyer. *Arch. Psychol.*, 1960, 37, 311-413.

WOHLWILL, J. F. Developmental studies of perception. *Psychol. Bull.*, 1960, 57, 249-288.

WOHLWILL, J. F. Some characteristics of psychophysical data obtained with a group version of the up-and-down method. *Percept. & Psychophys.*, 1968, 3, 54-56.

NOTE

1. Address: Department of Psychology, Michigan State University, East Lansing, Michigan.