Notes and Comment

Scaling loudness over short ranges: A reply to Poulton

ROBERT TEGHTSOONIAN
and MARTHA TEGHTSOONIAN
Smith College, Northampton, Massachusetts

Poulton (1984) has recently reopened an old debate about the form of the function relating loudness to the sound-pressure level of the signal. Despite the large body of evidence developed on this subject by S. S. Stevens and his associates showing that a power function provides a good description of loudness judgment both for tones and, to a less exact degree, for noise, Poulton now raises again the Fechnerian idea that loudness is proportional to the logarithm of sound intensity or, in the terms of his title, that loudness judgments are linear with decibels. In our view, however, the evidence offered to support this claim is unconvincing.

Two sets of data are presented. One set, published for the first time, concerns numerical judgments of loudness for a set of six intensities of white noise ranging from 80 to 85 dB(A) in steps of 1 dB. After two repetitions, intended to familiarize the subjects with the stimulus set, each stimulus was presented following the 80-dB stimulus, designated as the standard and assigned the value of 10 on the scale of loudness. The set was presented to all subjects in the same fixed irregular order three times in succession.

Before considering the results Poulton obtained, it is useful to consider what outcomes are to be expected given both the power law and the logarithmic rule proposed by Poulton. First, assume a power function with an exponent of 0.7 (a value consistent with both Poulton's data and those reported by R. Teghtsoonian and M. Teghtsoonian [1978] for the same range of intensities, but using 3-kHz tones). If the standard is called 10, the succeeding five should be called 10.84, 11.75, 12.74, 13.80, and 14.96, and since, as Baird and Noma (1975), as well as Poulton and others, have noted, unpracticed subjects are inclined to restrict themselves to integral values in the region of numbers near 10, the expected judgments for the full set of six are simply 10, 11, 12, 13, 14, and 15. By comparison, Poulton's proposal that the average subject adds one unit of loudness for every decibel means that beginning with a standard called 10, the remaining values will be called 11, 12, 13, 14, and 15. It belabors the obvious to note that the two predictions are identical. The numerical example simply makes concrete a well-

known principle, namely that the difference between a logarithmic relation and a power relation can be so small when examined over such a short range that the prospects of empirically differentiating between them approaches nil.

Nonetheless, Poulton presents, in his Figure 1, the median judgments obtained in each of the three successive presentations plotted against decibels shown on a linear scale. In Panel A, the judgments are shown on a linear ordinate, and in Panel B, the same values are shown on a logarithmic ordinate. The question is whether the data of Panel A or of Panel B are better fitted by straight lines. Not surprisingly, there is virtually no difference in the visually judged goodness of fit, but Poulton, in an act of inferential derring-do, declares that the third judgments show some concavity when plotted logarithmically and that, therefore, "the relation between loudness and intensity is linear in decibels." We reproduce Figure 1 here so that the reader can judge for himself the merits of the case. It seems to us that Poulton's conclusion is extravagant, and that he is on far better ground when, in the caption to Figure 2, showing plots for all 37 individual subjects, he states accurately that "Apart from functions 19 through 22, all the functions appear to be fitted as well by a straight line as by a simple curve." In that statement, he acknowledges the impossibility of preferring one model over the other on the basis of these data alone, but his general conclusion remains unaltered.

A second line of argument developed by Poulton is based on a previously reported study (Poulton, 1982). Data are reported for four large groups (ranging in size from 28 to 33) showing distributions for loudness judgments of the first stimuli following a standard noise at 65 dB SPL, which was assigned a value of 1 on the scale of loudness; that first stimulus was 70 dB for one group, 85 dB for two groups, and 100 dB for the fourth group. Median judgments for those four groups are reported as 2, 6, 7, and 20, respectively. Poulton construes those results as further evidence that loudness is linear with decibels, and makes his case in the following way. Assuming the linearity hypothesis to be true, the two 20-dB conditions are in fairly good agreement with each other, implying .25 loudness units per decibel in one case and .30 units per decibel in the other. The 5-dB case, too, is seen to be consistent with decibel linearity, although here the unit falls to .20. The 35-dB case appears to suggest a significant departure from the rule, since here, again assuming linearity, the unit rises to 0.54 (19 units divided by 35 dB). This, Poulton argues, is because the median judgment ought to be near 11 but is drawn to 20 because of a "logarithmic response bias."

In responding to such argument, one hardly knows where to begin. At the simplest level, it should be noted

The authors' mailing address is: Department of Psychology, Smith College, Northampton, MA 01063.

Figure 1. The least squares linear relation between judged loudness and the intensity of the noise in decibels. In Panel A, the median numerical responses of the 37 observers are plotted on a linear scale. On the third presentation, the function is well fitted by a straight line. In Panel B, the responses are plotted logarithmically. On the third presentation, the function is concave downward. (From Poulton, 1984.)

that when only two points are recorded, the standard and one comparison, no function can be ruled out. The 1982 data cited by Poulton can be said to support his linearity hypothesis in the same sense that they support any other hypothesis: They do not rule it out. But Poulton goes a step further and invokes a principle of consistency: If the linearity principle operates, he reasons, it ought to entail a loudness unit of approximately the same size in all four groups, and he is at some pains to make the 35-dB case conform to what is implied by the other three groups. But there seems to be little ground to his claim that, predicting from the other three groups, "the variable 35 dB above the standard . . . should receive a median judgment of 11." Following his reasoning, and assuming his linearity rule, those three groups generate predicted response values for the 35-dB case of 8.00, 9.75, and 11.50, respectively, and any reasonable interpretation of a number selection bias in favor of powers of 10 or integers less than 10 would predict the selection of 8, 9, or 10, not Poulton's predicted 11 or the obtained value of 20. In our view, these results conflict with, rather than support, the linearity rule.

We are also puzzled by the fact that, whereas in the 1984 study his rule calls for one unit per decibel, the 1982 study implies units that range from 0.20 to 0.54. If it is correct, as he argued, that there should be consistency in unit size among the four groups of the 1982 study, why isn't it also to be expected for the two separate studies? The group sizes in both studies are large and provide ample opportunity for the estimation of reliable average scores. It seems to us that the reason for a lack of consistency, both within the 1982 study and between it and the 1984 study, is the failure of Poulton's linearity

hypothesis to apply. Consider our previous argument that the results of Poulton's 1984 study, which he construes to imply a 1-unit-per-decibel rule, are also perfectly compatible with a power law. Although a choice cannot be made on the basis of that study alone, the wider ranges employed in the 1982 study do indeed allow a test of sharply different predictions. Predictions based on the unit-per-decibel rule are easy to develop. Starting with a standard called 1, values 5, 20, and 35 dB higher should be called 6, 21, and 36, respectively. Assuming a power law with an exponent of 0.7, the corresponding predicted values are 1.50, 5.01, and 16.8. We leave to the reader to decide which predictions agree better with the obtained values of 2.00, 6.00 (or 7.00), and 20.0.

At a more general level, we contend that it is probably a mistake to make claims, as Poulton appears to do, about what untrained subjects will do when given deliberately unstructured instructions to make number assignments to stimuli that are barely discriminable from one another. It seems likely to us that subjects will vary in how they respond, depending on how they understand the task and on how well they understand the nature of the number system. Some may do no more than discriminate between stimuli as best they can and achieve some rank ordering. For example, Poulton's Subject 35 called the two weakest signals 11 and the remaining four all 12. It is hard to take seriously any claim about such data other than the obvious implication that the subject formed two categories, one called "a little bit softer" and the other called "a little bit louder," and to humor the experimenter was prepared to use numbers as names for those categories. It also seems likely to us that some subjects may indeed, as Poulton argues, try to establish equal intervals and as-

sign numbers to reflect the boundaries of those intervals. And we feel sure that even under such restrictive conditions, there are some subjects who both see the task as one of judging sensation ratios and have a sufficiently sophisticated understanding of the system of numbers to use it for that purpose. Where such diversity exists, the problem becomes one of trying to determine which subjects are doing what with numbers. What then becomes the issue is not selection between alternative versions of a psychophysical law, but rather identification of the level of cognitive development attained by any given subject, or (even less interesting) the discovery of how ambiguous instructions are interpreted.

The conclusion we draw is that Poulton has brought the wrong method to the wrong situation and, as a result, has arrived at the wrong answers. Given the various interpretations to which his instructions can be put, and given the variety of levels at which subjects may understand and use the number system, and, particularly, given a very short range of barely discriminable stimuli, the results must of necessity be ambiguous, and the fact that they have not seemed so to Poulton is more testimony to the strength of his theoretical desires than to the validity of his argument. The question he raises is, of course, worth pursuing, but in our view that is best done by supplementing number-matching with cross-modal matching, where number-selection biases, and ambiguities about the ways in which numbers can be used, are avoided.

In summary, we believe that Poulton's claim to have shown that "loudness grows linearly with decibels" derives no support from the evidence he cites. First, the experiment he reports concerning loudness judgments over a 5-dB range cannot discriminate between a power law and a logarithmic law, and Poulton's contention that a visual inspection of the data favors the latter must be regarded as wishful thinking. Second, his claim, that the logarithmic law is supported by data in his 1982 report on first judgments of a comparison 5, 20, or 35 dB above a standard called 1.0, rests on his special plea that the 35-dB comparison "should have been called 11" when, in fact, the median (and modal) judgment was 20. Contrary to Poulton's interpretation, the median judgments for these three comparison levels fall remarkably close to values predicted by a power law. Third, we are skeptical about the value of a strategy that seeks to resolve an important theoretical issue by giving untrained observers deliberately ambiguous instructions about the performance of an extremely difficult task. Finally, we think it a serious misjudgment to urge, as Poulton appears to do, a major revision of an important psychological law on the basis of a reexamination of two isolated experiments and an almost total disregard for disconfirming evidence. Fundamental principles are always subject to revision and replacement, but not on such tenuous grounds as Poulton proposes.

REFERENCES

- BAIRD, J. C., & NOMA, E. (1975). Psychophysical study of numbers: I. Generation of numerical responses. *Psychological Research*, *37*, 281-297.
- POULTON, E. C. (1982). Biases in quantitative judgments. *Applied Ergonomics*, *13*, 31-42.
- POULTON, E. C. (1984). A linear relation between loudness and decibels. *Perception & Psychophysics*, *36*, 338-342.
- TEGHTSOONIAN, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, *78*, 71-80.
- TEGHTSOONIAN, R. (1973). Range effects in psychophysical scaling and a revision of Stevens' law. *American Journal of Psychology*, *86*, 3-27.
- TEGHTSOONIAN, R., & TEGHTSOONIAN, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, *24*, 305-314.

APPENDIX ON PREVIOUS RESEARCH

To set the record straight, we feel obliged to correct some mistaken impressions of our previously published research that may result from Poulton's account of it. The only experiment that we have reported using a 5-dB range of sound intensity was first cited in R. Teghtsoonian (1973) and later described in greater detail in R. Teghtsoonian and M. Teghtsoonian (1978). Briefly, the purpose of that condition was not to appraise the accuracy of the power law, but, in conjunction with companion groups spanning ranges of 10, 20, and 40 dB, to determine the effects of range on power law exponents. At issue was the question of whether the effects of range change for any single perceptual attribute (such as loudness) were as marked as those reported for range changes associated with the natural variation across a diverse set of perceptual attributes. R. Teghtsoonian (1971) reported evidence, based on a large group of scaled continua, that exponents were inversely proportional to stimulus range (in log units), a finding consistent with the view that widely varying dynamic ranges were all subjectively equal. It was critical to know whether the same result occurred when range was varied within a single perceptual attribute, since this finding would imply that magnitude-estimation subjects simply produce some fixed range of numbers no matter what range of intensities are presented, an outcome that would make ME useless as a psychophysical method. The important finding was that, although smaller ranges do yield larger exponents, the rate of growth is far smaller than the inverse proportionality noted for the cross-modality case. Poulton's argument that this reduction constitutes a flaw or at the least an unexpected outcome fails to recognize the important distinction we draw between the two types of range variation.

A few errors of fact and implication also require correction. We employed 16 subjects per condition, not 12 as Poulton asserts. We employed tones, not white noise, as Poulton's disclaimer for the relevance of his argument to the case of tones implies. We did, indeed, draw our subjects' attention to the ratios among the loudnesses they would experience, contrary to Poulton's assertion that we did not. And, although it is true that our rejection criterion for admitting subjects was a very modest value of 0.80 for the coefficient of determination, the overall average for all of our magnitude conditions was 0.92 and for the 5-dB condition was 0.90, indicating a rather stronger basis for the adequacy of the power law to describe our data than is implied by the figure of 0.80.