LETTERS TO THE EDITOR

Rating Scales in Detection Experiments

W. D. Larkin

*U. S. Army Personnel Research Office, Washington, D. C.**

(Received 7 December 1964)

The use of rating-scale isosensitivity data to test a discrete, two-state model for binary-choice detection experiments is shown to be inappropriate unless new assumptions are made. It is argued that, whatever assumptions are added, rating-scale isosensitivity curves will be different from binary-choice isosensitivity curves. The suggestion is made that a multistate model may be more appropriate than a two-state model for rating experiments, but the resulting isosensitivity curves appear to be too similar to those of signal-detectability theory to make rating-scale data useful for deciding between the theories.

Recently, papers have appeared in which rating scales are used to test psychophysical models of detection.[1–6] In this Letter, we take issue with a number of these papers, *viz.*, those in which the rating-scale data are cited as tests of threshold theories.

A rating scale is a set of ordered response alternatives $\{r_i\}$, $i=1,\cdots,k$, with which an observer indicates his degree of confidence that a stimulus event occurred. Denoting a stimulus occurrence by S and its nonoccurrence by N, the rating scale data are summarized by plotting

$$P(r\geqq x|\mathrm{S})=\sum_{i\geqq k-x}^{k} p(r_i|\mathrm{S}), \tag{1}$$

as a function of

$$P(r\geqq x|\mathrm{N})=\sum_{i\geqq k-x}^{k} p(r_i|\mathrm{N}), \tag{2}$$

for each $x=0,\cdots,k-1$, where $p(r_i|\cdot)$ is the proportion of responses in the ith rating category, conditional on the stimulus presented. This relation is analogous to the conventional isosensitivity curve from a "Yes–No" detection experiment.[7] Much of the appeal of the rating-scale experiment seems to be that isosensitivity data can be generated in a single session with fixed experimental conditions, whereas the Yes–No method combines data from several different experiments.

The similarity of the way in which the data are plotted seems to have been taken by some as sufficient evidence for the equivalence of the binary- and multiple-response designs, insofar as they bear on psychophysical theories. The theories in question fall into two classes: those that represent the effects of stimuli as continuous variables, e.g., signal-detectability theory,[4] and those that propose a discrete representation, e.g., neural quantum theory.[8] In a continuous theory, it is natural to assume that the observer can grade his detection observations in k steps, where k is the number of alternatives on the rating scale, and that, if we change to a Yes–No experiment, he can choose any criterion c such that, for an observation which would lead to a rating r_i, the response rule

If $i\geqq c$, respond YES
If $i<c$, respond NO

determines his binary choice. For each possible c, the equations

$$P_c(\text{YES}|\mathrm{S})=\sum_{i=c}^{k} p(r_i|\mathrm{S}) \tag{3}$$

and

$$P_c(\text{YES}|\mathrm{N})=\sum_{i=c}^{k} p(r_i|\mathrm{N}) \tag{4}$$

connect the rating scale with the binary-choice data.

The connection is not so easy to make if the underlying theory is discrete, rather than continuous. According to a threshold theory proposed by Luce[10] for Yes–No detection experiments, only two observation states are used—one when the threshold is passed and the other when it is not passed. The isosensitivity curve for this model is a pair of straight-line segments, one beginning at $\langle 0,0\rangle$ and the other at $\langle 1,1\rangle$, with a common endpoint in the interior of the unit square. Some authors[1,5] have used rating-scale isosensitivity data to test this prediction. Such comparisons are clearly inappropriate because, if an observer is required to choose among more than two response alternatives—as he must on a rating scale—new assumptions are needed to predict the response pattern.

A plausible argument can be made that, whatever new ones are added, only counterintuitive assumptions could make the two-state theory predict the same isosensitivity curve for rating experiments as it does for binary choices. Consider a continuous-response interval, such as the one Watson *et al.*[5] used, on which the distance x from one endpoint measures the observer's confidence rating. Let $F(x)=P(r\geqq x|\mathrm{S})$ and $G(x)=P(r\geqq x|\mathrm{N})$ be the two conditional probability distributions that govern the observer's ratings. A theory of the isosensitivity curve describes the relationship between F and G for every choice of x. If the theory predicts the same straight-line segments as Luce's[10] Yes–No theory does, the rating distributions would obey

$$F(x)=\alpha G(x),\quad 0\leqq G(x)\leqq a;\quad F(x)=\alpha' G(x)+\beta,\quad a<G(x)\leqq 1, \tag{5}$$

where $0\leqq F(x)$, $G(x)\leqq 1$, $0<\alpha'<1<\alpha$, and $\beta>0$. Setting $G(x_a)=a$ and $F(x_a)=b$, the intersection of the straight lines is at $\langle a,b\rangle$, which we assume is somewhere in the interior of the unit square, and above the line $F(x)=G(x)$. It is not difficult to prove two consequences of these restrictions that are obvious from the geometry: (1) wherever the corresponding density functions $f(x)$ and $g(x)$ exist, their derivatives have the same sign, and (2) at least one of the density functions must have a point of discontinuity. These constraints eliminate some otherwise likely response distributions, such as the examples shown in Fig. 1.

The discontinuity requirement is untestable if a response continuum is not used, but the restriction on the slopes cannot be avoided: it violates one's expectation that confidence in an event, as it is reflected in the responses over a rating scale, ought to be inversely related to lack of confidence in the event. Because both conditions depend critically on the behavior of the isosensitivity curve at a single "corner" point, their utility in the analysis of data is questionable. Nonetheless, from the point of view of theory, such response distributions can only be considered implausible: they suggest that, however the two-state model is extended for rating scales, the isosensitivity curve will be different from the prediction for binary-choice experiments.

To account for rating-scale data with a discrete theory of the sensory process, one cannot escape the possibility that an observer

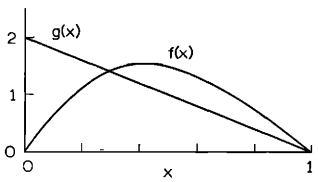

FIG. 1. Hypothetical response probability distributions over a continuous-rating interval. The continuity and the inversely related slopes that these curves exhibit must be disallowed if the isosensitivity relation is to conform to the two-state model. The examples shown are $g(x) = -2x+2$ and $f(x) = 4x(x-1) \times (x-2)$. These functions conform to the "smooth" isosensitivity curve $F(x) = G(x) + G(x) \times [1-G(x)]$, where $F(x)$ and $G(x)$ are the right-hand tails of the respective cumulative distributions.

may use more response alternatives than the number of sensory states. Nachmias and Steinman[3] dealt with this problem by assuming that, if the two-state model is true, the observer groups the alternatives into two equivalence classes. For purposes of testing the theory, these classes are indistinguishable from the binary choices "Yes" and "No," except that the class boundary is an additional free parameter.

A more likely possibility is that the two-state model is incorrect for rating experiments. This intuition is supported by several studies[11–13] that have shown that a discrete theory often requires more than two sensory states to account for detection data. In fact, Luce[10] went to more than two states to account for changes in response biases on psychometric functions. Support can also be found in the information analysis Watson *et al.*[5] report for their continuous-response data. Accordingly, we would expect the observer to make use of as many states as the number of responses permit. The result is a k-state model. If we denote the basic probability that the ith state is excited by a stimulus occurrence or nonoccurrence by $q_S(i)$ and $q_N(i)$, respectively, and assume unbiased responding, the k-state model asserts that, for each i, $P(r_i|S) = q_S(i)$ and $P(r_i|N) = q_N(i)$. This model predicts the coordinates of an array of $k-1$ points along the rating-scale isosensitivity curve:

$$\langle P(r \geqq x|N), P(r \geqq x|S)\rangle = \langle \sum_{i \geqq k-x}^{k} q_N(i), \sum_{i \geqq k-x}^{k} q_S(i)\rangle. \tag{6}$$

These points run convexly in the unit square from $\langle 0,0\rangle$ to $\langle 1,1\rangle$.

To trace out a full isosensitivity curve, we must introduce response biases, which we presume depend on, among other things, the instructions and the stimulus-event probability. In general, nothing stops the observer from giving rating i when his sensory state is j: he may in fact, choose, for each i, any k numbers $t_1, t_2, \cdots, t_k$, $0 < t_j < 1$, such that

$$P(r_i|S) = \sum_{j=1}^{k} t_j q_S(j), \tag{7}$$

$$P(r_i|N) = \sum_{j=1}^{k} t_j q_N(j), \tag{8}$$

subject only to the restriction that the left-hand sides of Eqs. (7) and (8) sum to unity. Without assumptions to limit these response "mixtures," we cannot predict the shape of the rating-scale isosensitivity curve, except that it must pass through the $k-1$ interior points of Eq. (6). We may anticipate that, even with a small number of states, any plausible limiting assumptions are likely to lead to curves that are too similar to those of a continuous theory to be distinguishable in data. This is true when we adopt a simple extension of Luce's idea that the observer biases from "Yes" to "No" (or the reverse) by adding a constant proportion of one response probability to the other. On the rating scale, we suppose that the observer chooses a constant t, $0 < t < 1$, and that he can bias his responses in two directions, toward higher ratings:

$$\begin{aligned} &P(r_1|S) = tq_S(1)\\ &P(r_i|S) = (1-t)q_S(i-1) + tq_S(i), \quad 1 < i < k,\\ &P(r_k|S) = (1-t)q_S(k-1) + q_S(k) \end{aligned} \tag{9}$$

or toward lower ratings:

$$\begin{aligned} &P(r_1|S) = (1-t)q_S(2) + q_S(1)\\ &P(r_i|S) = (1-t)q_S(i+1) + tq_S(i), \quad 1 < i < k,\\ &P(r_k|S) = tq_S(k). \end{aligned} \tag{10}$$

The equations for $P(r_i|N)$ are the same, except that q_N replaces q_S. Under either bias tendency, the isosensitivity curve is a piecewise linear function, ending at $\langle 0,0\rangle$ and $\langle 1,1\rangle$, with "corners" on the points of zero bias [Eq. (6)].

In summary, we find that some authors have drawn conclusions about a discrete threshold theory from data to which it does not apply. We have shown that a theoretical account of rating-scale responses cannot be expected to predict the same isosensitivity relation for rating scales as the two-state discrete theory predicts for Yes–No experiments. If the two-state theory is true, then a reasonable approach is to collapse the rating-scale alternatives into two classes and treat the data as though they came from a Yes–No experiment. However, a k-state representation may be more appropriate, in which case, without a theory of response bias, only $k-1$ isosensitivity points are predictable. These points are so arranged that, whatever bias assumptions are added to connect them, the multistate theory will be difficult to distinguish, in data, from a continuous theory.

Acknowledgments: I want to thank R. Duncan Luce and Robert T. Ollman for their help in preparing the manuscript.

* The opinions expressed are those of the author, and do not necessarily reflect official U. S. Army policy.

1 D. E. Broadbent and M. Gregory, "Vigilance Considered as a Statistical Decision," Brit. J. Psychol. **54**, 309–323 (1963).

2 J. P. Egan, A. I. Schulman, and G. Z. Greenberg, "Operating Characteristics Determined by Binary Decisions and by Ratings," J. Acoust. Soc. Am. **31**, 768–773 (1959).

3 J. Nachmias and R. M. Steinman, "Study of Absolute Visual Detection by the Rating Scale Method," J. Opt. Soc. Am. **53**, 1206–1213 (1963).

4 J. A. Swets, W. P. Tanner, and T. G. Birdsall, "Decision Processes in Perception," Psychol. Rev. **68**, 301–340 (1961).

5 C. S. Watson, M. E. Rilling, and W. T. Bourbon, "Receiver-Operating Characteristics Determined by a Mechanical Analog to the Rating Scale," J. Acoust. Soc. Am. **36**, 283–288 (1964).

6 D. J. Weintraub and H. W. Hake, "Visual Discrimination, an Interpretation in Terms of Detectability Theory," J. Opt. Soc. Am. **52**, 1179–1184 (1962).

7 An isosensitivity curve (or receiver operating characteristic) is, conventionally, the relation between $P(\text{YES}|S)$ and $P(\text{YES}|N)$, which is traced out for a series of binary-choice (Yes–No) detection experiments with fixed stimulus parameters and varying payoffs or instructions.

8 R. D. Luce, "Detection and Recognition," in *Handbook of Mathematical Psychology*, R. D. Luce, R. R. Bush, and E. Galanter, Eds. (John Wiley & Sons, Inc., New York, 1963), Vol. 1, p. 116.

9 We adopt the convention that, if $i > j$, r_i reflects a "stronger" observation than r_j, and, therefore, the observer is "more confident" that a stimulus event occurred when his rating is r_i than when it is r_j.

10 R. D. Luce, "A Threshold Theory for Simple Detection Experiments," Psychol. Rev. **70**, 61–79 (1963).

11 W. D. Larkin and D. A. Norman, "An Extension and Experimental Analysis of the Neural Quantum Theory," in *Studies in Mathematical Psychology*, R. C. Atkinson, Ed. (Stanford University Press, Stanford, Calif., 1964), p. 188.

12 D. A. Norman, "Sensory Thresholds and Response Bias," J. Acoust. Soc. Am. **35**, 1432–1441 (1963).

13 D. A. Norman, "Sensory Thresholds, Response Biases, and the Neural Quantum Theory," J. Math. Psychol. **1**, 88–120 (1964).

Erratum: Central Periodicity Pitch

[J. Acoust. Soc. Am. 37, 136–138 (1965)]

PHILIP C. NIEDER AND C. DOUGLAS CREELMAN

Research Center, Subcommittee on Noise, Los Angeles, California

(Received 16 February 1965)

IN THE TRANSITION FROM GALLEY TO PAGE, THE PRINTERS MISplaced one line of this article. The line, that belongs next to the bottom of the first column of p. 137 is to be found seventh from the bottom of the second column on p. 137.