Representations for biased numerical judgments

ROBERT F. FAGOT

University of Oregon, Eugene, Oregon 97403

MANARD R. STEWART

Northern Illinois University, DeKalb, Illinois 60115

and

RICHARD E. KLEINKNECHT

University of South Carolina, Columbia, South Carolina 29208

Models are presented that introduce bias parameters into the numerical representations for ratio and interval judgments. An interval bias model is considered that has the capability of accounting for both interval and ratio judgments. A contrasting ratio bias model—a simple generalization of the classical model that accepts ratio judgments at face value—is also considered, and both models are tested with brightness data. The ratio estimation data lead to the conclusion that the constraints placed on the structure of the data by the classical model are too strong. In terms of goodness-of-fit and certain theoretical criteria, the interval bias model accounts for the ratio estimation data better than the ratio bias model. The interval estimation data are supportive of recent efforts directed toward the development of direct interval scaling theory.

As Ekman and Sjöberg (1965) point out, the application of scaling methods in psychophysics is dominated by the use of direct estimation methods that assume subjects are capable of judging sensory ratios and that purport to construct ratio scales from the subjects' responses. S. S. Stevens has been the major contributor to this view of psychophysical scaling theory, and undoubtedly much of the work that followed along these lines was stimulated by Stevens (1957) and Stevens and Galanter (1957).

Recent theoretical investigations (Krantz, 1972; Marley, 1972) incorporate several empirical generalizations of ratio scaling as at least good approximations that must be accounted for in any theory of ratio scaling. Yet several empirical studies (Eisler, 1960; Fagot & Stewart, 1969; Goode, 1962; Mashhour, Note 1; Sjöberg, Note 2, 1971; and Svenson & Akesson, Note 3, Note 4), in which tests of internal consistency have been carried out, raise serious questions concerning the validity of scales based on ratio estimation data, and hence of some of the empirical generalizations accepted by Krantz and Marley. These empirical studies reveal the presence of systematic errors in ratio estimation judgments, and suggest strongly that the simple model that accepts

subjects' responses at face value must be abandoned. One method of accounting for the apparent bias is to introduce additional parameters into the numerical representations, a step taken in the present paper and in some of the studies cited above (see especially Sjöberg, 1971).

Although, as noted above, most scaling research in psychophysics has focused on ratio methods, recent theoretical and empirical studies of interval scaling methods (Pfanzagl, 1968; Fagot & Stewart, 1970) encourage the development of scaling theory based on numerical estimation of intervals. Hence, in this paper, a theory based on *biased* numerical judgments of both ratios and intervals will be considered.

In the section on theory, an interval model incorporating a bias parameter is presented. A special feature of this model is that it accounts for both interval and ratio judgments. A contrasting ratio bias model, a simple generalization of the Stevens model and one investigated by Sjöberg (1971) and others, is also presented, and both models are tested with brightness data, as well as compared to the simple models that assume no bias.

THEORY

The theoretical treatment will begin with a bisection system and show that a generalization of the bisection numerical representation provides a theoretical framework that not only accounts for bisection and multisection operations, but also both interval and ratio numerical responses.

This research was supported by a grant from the Advanced Research Projects Agency of the Department of Defense, and was monitored by the Air Force Office of Scientific Research under Contract No. F44620-67-C-0099 to the University of Oregon. Preparation of the manuscript was carried out in part by Fagot during his tenure as a Fellow at the Netherlands Institute for Advanced Study, Wassenaar, Holland, 1972-73.

In bisection, the subject is presented with two stimuli, a and b , and his task is to perform an *operation* that produces a stimulus (aob) that bisects the interval between a and b . A numerical representation for bisection (Pfanzenagl, 1968) is given by the following equation:

$$\psi_{aob} = a\psi_a + (1-a)\psi_b \quad (0 < a < 1) \quad (1)$$

where ψ is a real-valued scaling function, and a may be interpreted as a response bias parameter (Fagot & Stewart, 1970).

A generalization of the bisection numerical representation for q -section operations may be formulated as follows:

$$\psi_{aoqb} = \delta(1-q)\psi_a + [1-\delta(1-q)]\psi_b. \quad (2)$$

$\delta = 1$ implies the absence of bias. Here we have decomposed the a of the first representation (Equation 1) into two components— δ , a bias parameter, and q , the degree of the sectioning operation [i.e., $a = \delta(1-q)$]. For the q -section operation method, the experimenter fixes q , and the subject's task would be to perform an operation that produces a stimulus ($aoqb$) that partitions the total interval $\psi_b - \psi_a$ into two intervals $\psi_b - \psi_{aoqb}$ and $\psi_{aoqb} - \psi_a$. The constant q is a function of the ratio of the *lower* interval to the total interval (q equals this ratio in the absence of bias). If the smaller interval is taken as a unit, then the larger of the two intervals is equal to $(1-q)/q$. For example, if the experimenter sets $q = 1/3$, the subject produces two intervals such that the smaller interval (which is the bottom interval) is $1/2$ the magnitude of the larger interval and $1/3$ the magnitude of the total interval. The subject's "productions" are then the observations, and Equation 2 would be applied to estimate scale values and the parameter δ .

With numerical response methods, the subject does not perform an operation, but instead assigns a number directly to a stimulus. Let $\Phi_i < \Phi_j < \Phi_k$ be physical scale values for three stimuli with corresponding psychological scale values $\psi_i < \psi_j < \psi_k$; and let e_{ijk} denote the subject's theoretical numerical response. Then solving Equation 2 for q and making the appropriate notational changes results in:

$$e_{ijk} = \frac{1}{\delta} \left[\frac{\psi_j - \psi_i}{\psi_k - \psi_i} \right] + (1 - 1/\delta), \quad (3)$$

or equivalently,

$$\frac{\psi_j - \psi_i}{\psi_k - \psi_i} = \delta e_{ijk} + (1 - \delta). \quad (4)$$

Equation 3 shows that the theoretical response e_{ijk} (of the ratio of the lower interval to the total interval) is a linear function of the corresponding ratio in scale values with slope $(1/\delta)$ and intercept $(1 - 1/\delta)$.¹

In an appropriate scaling experiment, the subject's task would be to rate the middle stimulus (Φ_j) relative to the two standards (Φ_i and Φ_k). Thus the subject would report an estimate—denoted r_{ijk} —of the ratio of the lower interval to the total interval; i.e., an estimate of e_{ijk} , the parameter value. Given estimates of e_{ijk} based on the observations r_{ijk} , the scale values and the parameter δ would be estimated.

A numerical representation for ratio judgments (i.e., ratios of magnitudes), in which the subject's responses are taken at face value (assumed to be unbiased), is given by Equation 5:

$$e_{ij} = \psi_i / \psi_j \quad (5)$$

Equation 5 will be referred to as the *classical model*, and abbreviated *C model*.

In a procedure with an operation (e.g., fractionation), e_{ij} would be a number fixed by the experimenter, and the subject would produce a stimulus Φ_i relative to another stimulus Φ_j such that Equation 5 is assumed to hold. In ratio estimation experiments, Φ_i (the comparison or variable stimulus) and Φ_j (the standard stimulus) would be the presentation pair, e_{ij} would be a theoretical numerical response, and r_{ij} would denote the subject's estimate of e_{ij} .

The numerical representation for the interval model (Equation 3) can be applied to ratio judgments by assuming that ψ_i partitions the interval between the *threshold* t and ψ_j into two intervals, $\psi_j - \psi_i$ and $\psi_i - \psi_t = \psi_i$ (since $\psi_t = 0$), and hence that a ratio judgment can be interpreted as an interval judgment. By application of Equation 3 to ratio judgments, we get the following equivalent equations for the representation:

$$e_{ij} = (1/\delta R) \frac{\psi_i}{\psi_j} + (1 - 1/\delta R) \quad (6)$$

$$\frac{\psi_i}{\psi_j} = \delta R e_{ij} + (1 - \delta R) \quad (7)$$

Since the bias is not necessarily the same for a presentation pair (ratio task) as for a presentation triple (interval task), the two bias parameters for the general interval model are distinguished by use of a subscript in Equations 6 and 7.

A numerical representation for the ratio bias model, the multiplicative ratio model—MR model—is given by

$$e_{ij} = \beta \psi_i / \psi_j. \quad (8)$$

In the absence of bias, $\beta = \delta_R = 1$, and the representations given by Equations 6 and 8 reduce to the C model (Equation 5).

The relationship between the two models of ratio judgments is further explicated by writing Equations 6 and 8 as follows:

$$1 - e_{ij} = (\psi_j - \psi_i) / \delta_R \psi_j \quad (9)$$

$$1 - e_{ij} = (\psi_j - \beta \psi_i) / \psi_j \quad (10)$$

Equation 10 shows that for the MR model as well, ratio judgments may be interpreted as interval judgments, although if bias is present, it is as if the comparison stimulus is weighted by the bias parameter before the interval is "constructed."

Note on Terminology

In spite of the fact that a ratio judgment may be interpreted as an interval judgment for the MR model, the representation given by Equation 6 will be referred to as the interval model—abbreviated I model—as applied to ratio judgments or the ratio task), since Equation 6 is a special case of the general interval model given by Equation 3. A judgment r_{ij} relative to a presentation pair will always be referred to as a *ratio judgment*, and reference will be made to the *ratio task*. A judgment r_{ijk} will, of course, always be referred to as an *interval judgment*, and reference will be made to the *interval task*.

Axiomatization

If one were to attempt to axiomatize the above theories, it would be necessary in each case to construct a set of axioms sufficient to guarantee the existence of the appropriate numerical representation (Suppes & Zinnes, 1963). One such axiom for the C model (Equation 5) is the following *product condition*:

$$e_{ik} = e_{ije_{jk}}, \quad (\psi_i < \psi_j < \psi_k) \quad (11)$$

a necessary condition that follows immediately from Equation 5. It can be shown that Equation 11 and two additional simple axioms which would not ordinarily be tested, comprise a set of axioms sufficient to guarantee the existence of the numerical representation given by Equation 5. Hence Equation 11 contains the principal directly testable consequences of the C model.

A comparable product condition based on interval judgments, derived from Equation 4 with $\delta = 1$, is given below:

$$e_{hik} = e_{hij} e_{hjk}, \quad (\psi_h < \psi_i < \psi_j < \psi_k) \quad (12)$$

Axioms comparable to Equations 11 and 12, expressed entirely in terms of observables, have not

yet been formulated for the more complex bias models, and hence the empirical evaluation of the bias models will be restricted to the numerical representations directly.

Related Theoretical Work

The C model has been applied by most investigators using direct ratio methods, and corresponds to Sjöberg's (Note 2) Model A. The MR model (but not under this name) was first proposed by Eisler (1960) and adopted as Model B by Sjöberg (Note 2), who also proposed a Model C that assigned a different scale value to a stimulus depending on its status as comparison (variable) or standard stimulus. The ratio between the two scale values for the stimulus was called the V/S ratio. The MR model can be viewed as a special case of Model C in which the V/S ratios are assumed constant for all stimuli. Model C has the disadvantage of an excessively large number of parameters, and will not be tested against the data in this paper.

What does appear to be original with the present formulation is the bias interval model (Equation 3), and of special interest is its applicability to ratio judgments (Equation 6).

METHOD

General

Subjects performed two tasks, an interval task and a ratio task, both with respect to brightness. On the interval task, the subjects made percent judgments of intervals of brightness; on the ratio task, the subjects judged the percent brightness of a comparison stimulus relative to a standard. Adaptation time was 15 min. Subjects were allowed to take as long as desired to make a judgment.

Subjects

In Experiment 1, the subjects were 8 undergraduates and 2 graduate students; in Experiment 2, the subjects were 10 undergraduates. All subjects were paid \$1.50/h. None of the subjects had had previous experience in making brightness judgments.

Apparatus

The apparatus has been described in detail by Eskildsen (1963). It provides four luminous disks of white light, each target 13 mm in diam, placed horizontally and separated by 90 mm, center to center. Two of the targets were used for the ratio task, and three of the targets for the interval task. The apparatus was located in a light-tight visual tunnel, and the subjects were seated at a distance of approximately 3.66 m from the targets in an ordinary classroom desk seat. An intercom system was used for oral presentation of the subject's response.

Procedure

Experiment 1: Five of the 10 subjects were randomly assigned to perform the interval task first and the ratio task second, and the remaining 5 subjects performed the tasks in reverse order. Experiment 2: Five of the subjects were assigned randomly to the interval task and the remaining 5 to the ratio task.

The instructions and procedure for each task were the same for both experiments as described below.

Interval Task

The instructions for the interval task were as follows: "In this experiment you will be required to judge the *brightness* of lights. On each trial of this part of the experiment you will be presented with three lights, which will vary from trial to trial. The light on the left will always be the dimmest of the three, the light on the extreme right will always be the brightest of the three, and the middle light will be intermediate in brightness, so the three lights will increase in brightness from left to right. You will be asked to make your judgments in terms of *intervals* of brightness, by rating the middle light as a percentage of the *distance* in brightness between the two outer stimuli. Now, let's look at this diagram on the blackboard to make sure you understand the way in which you should make your judgments."

A diagram on a blackboard was then used to illustrate the procedure and to insure that the subjects thoroughly understood how to make the interval judgments.

Six stimuli were selected and all possible 20 triples of stimuli were formed. A trial of the interval task consisted of the presentation of a triple and a percentage response from the subject. The percentage response (r) corresponded to a *ratio* of magnitude intervals. The set of 20 triples was replicated 3 times in each session, for a total of 60 observations per session. Each subject participated in three sessions, one each day for 3 days. A different random order was used for each set of 20 triples. A total of nine observations was made relative to each of the 20 triples.

Stimuli were selected to give a wide range of ratios of intervals. The actual stimuli used were .1, .2, 1, 9, 80, and 600 fL. Prior to the experiment, it was computed that use of these six stimuli should provide a wide range of ratios of intervals for the subjects with exponents (of a power function) around .10 or below, the expected range for subjects making interval judgments.

Ratio Task

The instructions for the ratio task were as follows: "In this experiment you will be required to judge the *brightness* of lights. On each trial of this part of the experiment you will be presented with two lights which will vary from trial to trial. The light on the right will always be the brighter of the two, and will be called the *standard*. Your task will be to judge the brightness of the dimmer light on the left *in relation* to the standard on the right. You should report the brightness of the dim light as a *percentage* of the brightness of the standard. For example, if you see the dim light as about half as bright as the standard, you should report its brightness as 50%, i.e., that it appears to you to be 50% as bright as the standard on the right. On the other hand, if the light appears about 1/10 as bright as the standard, you should respond 10%, whereas if the dim light appears about 3/4 as bright as the standard, you should report 75%; and so forth. Now, although the light on the left will always be dimmer than the standard on the right, its brightness in relation to the standard will be quite variable and you may use any percent greater than 0 and less than 100 that represents to you the brightness of the dim light in relation to the standard. Any questions?"

Ratio judgments derived from the first day were used only to compute, for each subject, the exponent of an assumed power function, which in turn was used to generate a set of stimuli which would provide a wide range of ratios for each subject. In order to construct a hypothetical wide range of ratios for each subject based on the power function exponent computed the first day, it was necessary to use a different set of stimuli for each subject. For all subjects, the highest stimulus (Φ_6) was set at 600 fL, and a typical set of values was 0.1, 1.5, 35, 180, 370, and 600 fL.

For each subject, the six stimuli were taken two at a time to form 15 presentation pairs. A single trial of the ratio task consisted of the presentation of one pair of stimuli and a percentage response from the subject. The set of 15 pairs was replicated 3 times in each session, for a total of 45 observations per session. Each subject participated in three sessions, one each day for 3 days. A different random order was used for each set of 15 pairs. A total of nine observations was made relative to each of the 15 pairs.

RESULTS AND DISCUSSION

The analysis of data consisted of a least squares procedure for estimating scale value and bias parameters; statistical tests of systematic trends; "improvement" tests, i.e., tests of the improvement in fit produced by introduction of a bias parameter into a given model; and an analysis of the effect of model and task on scale values.

All the model equations are in linear form which facilitates the application of regression procedures for estimating the parameters and testing hypotheses. However, iterative procedures were required for the simultaneous least squares estimation of bias parameters and scale values. The dependent variable was r_{ijk} for the interval task and r_{ij} for the ratio task.

The estimation procedure was carried out as follows: Let ψ be a vector of scale values. A response r_{ij} or r_{ijk} was assumed to be some function of the scale values and bias parameter, the particular function depending upon the model equation. Let SSD denote the total deviations sums of squares between observed responses and values predicted by the model equation. Then,

$$SSD(d) = m \sum_{i,j,k} [\bar{r}_{ijk} - f(\psi, d)]^2$$

$$SSD(d_R) = m \sum_{i,j} [\bar{r}_{ij} - g(\psi, d_R)]^2$$

$$SSD(\beta) = m \sum_{i,j} [\bar{r}_{ij} - h(\psi, \beta)]^2$$

where $m = 9$ replications, $\bar{r}$ denotes the mean of the 9 replications, and the functions f , g , and h are the right-hand sides of Equations 3, 6, and 8, respectively. For the no-bias models, the appropriate bias parameter was set equal to one, and hence only the scale values needed to be estimated. Using the Gauss-Newton least-squares method (Hartley, 1961), scale values and bias parameters were estimated such that the appropriate SSD was minimized. For a discussion of the distributional properties of these estimates, see Hemmerle (1967).

Using scale values and bias parameters estimated from all the data, predicted values $\hat{q}_{ij}$ and $\hat{q}_{ijk}$ were computed. For the interval task, estimates $\hat{q}_{ijk}$ were computed using Equation 3 for each of the 20 triples. On the ratio task, for each of the 15 pairs, estimates $\hat{q}_{ij}$ were computed using the I model (Equation 6) and the MR model (Equation 8). Estimates under the assumption of no bias were obtained by setting the appropriate parameter value equal to unity. The procedure outlined above was carried out for each subject separately.

Figure 1 shows a plot of observed $\bar{r}_{ijk}$ as a function of values $\hat{q}_{ijk}$ predicted by the interval bias model. Each subject contributed 20 data points.

Figure 1. Observed $\bar{r}_{ijk}$ as a function of values $\hat{e}_{ijk}$ predicted by the interval bias model.

Figure 2. Observed $\bar{r}_{ij}$ as a function of values $\hat{e}_{ij}$ predicted by the I model.

Figure 2 shows a plot of observed $\bar{r}_{ij}$ as a function of values $\hat{e}_{ij}$ predicted by the I model; Figure 3 shows a plot of observed $\bar{r}_{ij}$ as a function of values $\hat{e}_{ij}$ predicted by the MR model; and Figure 4 shows a plot of observed $\bar{r}_{ij}$ as a function of values $\hat{e}_{ij}$ predicted by the C model. Each subject contributed 15 data points in Figures 2, 3, and 4.

The value of the bias models (contrasted with the no-bias models) is not obvious from an inspection of Figures 1-4, especially since these plots are for all subjects combined. One approach to the question is to

test for improvement in fit produced by the introduction of a bias parameter into a given model. The first step was to compute deviations sums of squares (SSD) from the residuals $\bar{r}_{ijk} - \hat{e}_{ijk}$ (interval task) and $\bar{r}_{ij} - \hat{e}_{ij}$ (ratio task) for each model and subject separately, where the $\bar{r}_{ijk}$ and $\bar{r}_{ij}$ are the means of nine replications. The improvement test uses an "extra sum of squares" principle (Draper & Smith, 1966, pp. 67-71). It involves computing the difference between SSD for a given model with and without the bias parameter. The corresponding deviations mean

Figure 3. Observed $\bar{r}_{ij}$ as a function of values $\hat{e}_{ij}$ predicted by the MR model.

Figure 4. Observed $\bar{r}_{ij}$ as a function of values $\hat{e}_{ij}$ predicted by the C model.

Table 1
Bias Parameter Estimates and Improvement Tests: Interval Task

Subject	MSQE	SSD ($\delta = 1$)	SSD ($\hat{\delta}$)	Percentage SSD Reduction	F†	$\hat{\delta}$
IR						
K.H.	.0097	.1500	.1099	26.7	4.13*	1.066
D.S.	.0071	.1618	.1555	3.9	.88	1.023
M.W.	.0045	.0928	.0755	18.6	3.84	1.037
W.H.	.0070	1.5461	.8677	43.7	96.06**	.770
D.L.	.0041	.2026	.1943	4.1	2.02	.974
RI						
L.B.	.0156	1.2794	.9149	28.5	23.37**	.840
J.B.	.0091	.8523	.8518	.1	.05	1.004
L.C.	.0036	.1267	.1183	6.6	2.33	1.017
D.S.	.0055	1.0802	.9214	14.7	28.87**	.885
D.K.	.0071	.3307	.2194	33.7	15.68**	1.068
I						
G.F.	.0074	.9032	.5457	39.6	48.31**	1.197
H.D.	.0104	.4043	.3943	2.5	.96	1.030
K.M.	.0097	.3930	.3706	5.7	2.31	.962
L.S.	.0114	.2289	.2286	.1	.03	1.005
A.M.	.0090	.7690	.7255	5.7	4.83*	.943

Note—IR = performed interval task first, ratio task second, RI = performed ratio task first, interval task second, I = performed interval task only.

†Improvement test based on 1 and 160 df.

* $p < .05$

** $p < .001$

square (MSQD) is then tested against the appropriate error mean square (MSQE). The F tests of significance were based on a subject's own MSQE; a pooled MSQE was not used since $F_{\max}$ tests resulted in rejection of the hypothesis of equality of error variance across subjects for both tasks.

Using this procedure, each of the three bias models was compared with its appropriate no-bias model to test for possible significant improvement in fit due to the addition of the appropriate bias parameter. This

technique has been used by Sjöberg (1971) to compare certain of the ratio estimation models used in the present analysis.

Table 1 presents bias parameter estimates and results of improvement tests for the interval task, and Table 2 presents this same information for the ratio task.

Note that on the interval task, 7 of the 15 subjects showed significant improvement through use of the bias parameter. On the ratio task, 10 of the 15

Table 2
Bias Parameter Estimates and Improvement Tests: Ratio Task

Subject	MR Model						I Model			
	SSD		Percentage SS		$\hat{\beta}$	Percentage SS		$\hat{\delta}_R$		
	MSQE ($\delta_R = \beta = 1$)	SSD	Reduction	F†		SSD	Reduction		F†	
IR										
K.H.	.0063	.0997	.0989	.8	.13	1.010	.0378	62.1	9.83*	.872
D.S.	.0062	.0584	.0544	6.8	.65	.971	.0577	1.2	.11	.963
M.W.	.0023	.9488	.7702	18.8	77.65**	1.146	.4073	57.1	235.43**	.217
W.H.	.0054	1.1543	.8826	23.5	50.31**	1.192	.6857	40.6	86.78**	.578
D.L.	.0062	.3012	.1647	45.3	22.02**	1.129	.0612	79.7	38.71**	.508
RI										
L.B.	.0123	.8321	.5895	29.2	19.72**	1.163	.4884	41.3	27.94**	.520
J.B.	.0069	.0807	.0807	.0	.00	1.003	.0803	.5	.06	.990
L.C.	.0059	.4093	.1946	52.5	36.39**	1.177	.0484	.2	.15	1.014
D.S.	.0024	.5622	.4234	24.7	57.83**	1.132	.4058	27.8	65.17**	.719
D.K.	.0115	.0602	.0600	.3	.02	.994	.5069	5.5	.29	.966
R										
T.M.	.0111	.1445	.1444	.1	.01	.999	.1430	1.0	.14	.982
D.R.	.0056	.4744	.2828	40.4	34.21**	1.135	.2841	40.1	33.98**	.515
T.T.	.0155	.3057	.1472	51.8	10.23*	1.132	.0940	69.3	13.66**	.570
G.S.	.0088	.9025	.6413	28.9	29.68**	1.178	.6164	31.7	32.51**	.550
R.H.	.0060	.2356	.1849	21.5	8.45*	1.074	.0711	69.8	27.42**	.717

†Improvement tests based on 1 and 120 df.

* $p < .01$

** $p < .001$

Note—IR, RI: See Table 1. R = performed ratio task only.

subjects showed significant improvement for each of the two models, 9 of these 10 subjects showing significant improvement on both models.

Evidence that the I model is providing a better fit to ratio task data than the MR model is given by the fact that an ordinal comparison of the relative magnitude of SSD by the subject favors the I model: 12 of the 15 subjects who performed the ratio task had smaller SSD for the I model. Using the Wilcoxon test for two matched samples (or the same subjects under two conditions), the outcome is that the observed proportion (0.80) is significantly different from .50 ($Z = 2.22$, $p = .028$, two-tailed test). The same kind of analysis of interval and ratio task data (based on 10 common subjects) resulted in no significant differences either for MSQD or MSQE.

Tables 1 and 2 also show the percent reduction in SSD through use of a given bias parameter for each subject separately. For the seven subjects who showed significant improvement on the interval task, the mean reduction was 27.5%, with a range of 5.7% to 43.7%.

On the ratio task, the 10 subjects who showed significant improvement using the MR model had a mean of 33.7%, with a range of 18.8% to 52.5%, and the 10 subjects who showed significant improvement using the I model had a mean of 51.5%, with a range of 27.8% to 79.7%. Eight of the 9 subjects who showed significant improvement for both the MR and the I model had a larger reduction in SSD using the I model, a result significant at the .05 level (binomial, two-tailed test). Hence, most subjects showed smaller SSD for the I model, and of those subjects who showed significant improvement through use of the

Figure 5. Fit of the ratio product condition (Equation 11): observed $\bar{r}_{ik}$ as a function of predicted values $\bar{r}_{ij}\bar{r}_{jk}$.

Figure 6. Fit of the interval product condition (Equation 12): observed $\bar{r}_{hik}$ as a function of predicted values $\bar{r}_{hij}\bar{r}_{hjk}$. The large circle near the origin represents 12 points.

bias parameter, the improvement was greater for most subjects for the I model.

The above analysis addresses itself to goodness of fit, but another important question is the possible presence of systematic trends in Figures 1-4. Inspection of Figure 1 indicates a reasonably good overall fit for interval task data, with no obvious systematic departures except possibly a tendency for predicted values to be too low relative to observed values for very high ratios. For ratio task data—Figures 2, 3, and 4—there is the suggestion of a systematic trend: a tendency toward an S-shaped function, somewhat more noticeable for the C model.

The improvement test was used to test for the presence of a cubic trend in Figures 1-4, as well as for corresponding figures for individual subjects. It involved computing the reduction in SSD obtained by adding the cubic term to the regression model. The corresponding MSQD was then tested against the appropriate MSQE (as given in Tables 1 and 2 for individual subjects). This procedure was carried out separately for each subject, for all subjects combined, and for all models. No significant cubic trend was found for any individual subject for any model. The only significant cubic trend was for all subjects combined for the C model (Figure 4), $F(1,1800) = 5.53$, $p < .025$.

Product Conditions

Figure 5 shows a plot of observed $\bar{r}_{ik}$ as a function of predicted values $\bar{r}_{ij}\bar{r}_{jk}$, and provides a graphical display of the fit of the ratio product condition (Equation 11) to the ratio task data.

Figure 6 exhibits the fit of interval task data to the interval product condition (Equation 12), and shows a plot of observed $\bar{r}_{hik}$ as a function of predicted values $\bar{r}_{hij}\bar{r}_{hjk}$.

Figure 7. I-model and MR-model scale values as a function of C-model scale values for 10 subjects who showed significant improvement on improvement test.

Figure 5 suggests an extreme systematic trend, with predicted values generally much too high relative to observed values for moderate and low ratios (below about 0.60). On the other hand, there is no obvious systematic trend in the interval task data (Figure 6).

Two different methods were used to test for the presence of systematic trends in the product condition data. In the first method, each subject was assigned a score defined as the proportion of predicted values larger than observed values (corresponding to the proportion of points below the identity line). For each product condition, a t test of the hypothesis $H_0: \mu = .50$ was carried out with a sample size $n = 15$. For ratio task data, the null hypothesis was rejected ($t = 2.64$, $p < .02$, two-tailed test) and in 77.7% of the cases predicted values were larger than observed values (for all subjects combined). For interval task data, the null hypothesis was not rejected ($t = 0.48$), and in only 46.2% of the cases were the predicted values larger than the observed values.

The statistically significant result obtained for the ratio task data would be more meaningful if a significant number of subjects showed the same directional trend. A count showed that for 13 of the 15 subjects on the ratio task, the trend was in the same direction (greater than 50% of the predicted values higher than appropriate observed values). Using the binomial, this result is significant for a two-tailed test ($p < .01$). For interval task data, the result was not significant, with 7 of 15 subjects having a majority of predicted values higher than corresponding observed values.

Another issue of consequence is the effect on scale value estimates of model and task. We consider first the effect of changes brought about by use of bias parameters.

Effect of Model on Scale Values

For the ratio task, Figure 7 shows a plot of scale value estimates generated by the bias models as a function of scale values estimated by the C model.

The diagonal is the theoretical line for the hypothetical case in which the C model holds exactly. The points above the diagonal represent a plot of I-model estimates as a function of C-model estimates, and the points below the diagonal represent a plot of MR-model estimates as a function of the C-model estimates. Only the 10 subjects who showed significant improvement on the improvement test have been plotted. The data points for the five subjects who showed no significant improvement for each model lie extremely close to the diagonal, as can be inferred from the fact that the bias parameter estimates are very close to unity (see Table 2).

Note that the two models have opposite directional effects on the magnitude of C-model estimates: The I-model estimates are higher and the MR-model estimates lower than corresponding C-model estimates. In addition, the effect of the bias parameter is very much greater for the I model except for very high values. For example, scale value estimates for the dimmest stimulus (ψ_1) show the I-model estimates range (across the 10 subjects) from 1.72 to 4.10 times as large as C-model estimates (with a mean of 2.90), whereas C-model estimates range from 1.18 to 1.56 times as large as MR-model estimates, with a mean of 1.39 (note the change of direction relative to C-model estimates). The data points for these estimates all lie between about 10 and 22 on the horizontal axis in Figure 7.

The apparent convergence of data points toward the diagonal from the low to the high end of the scales for the I model is produced because estimates cannot exceed 100 (assigned to the brightest stimulus) but must lie above the diagonal (since bias parameter estimates are below unity for the I model). No such constraint is placed on the MR-model data since bias estimates are above unity.

The fact that in Figure 7 all the data points for the I model are above the diagonal ($\delta_R < 1$) implies (a) $e_{ij} < \psi_i / \psi_j$; and that the data points for the MR model are below the diagonal ($\beta > 1$) implies (b) $e_{ij} > \psi_i / \psi_j$. It is tempting to conclude that Case a defines a *contrast effect* and Case b an *assimilation effect*. However, such a conclusion assumes that the bias is due to a *relative size* factor. But as Sjöberg (1971) points out, if the standard is assigned a fixed number (as in the present experiment) then bias may result also from a so-called "*fixed-free*" factor. Sjöberg concluded that both factors were presented in his experiment on lifted weights, with $\beta > 1$ implying the presence of an assimilation effect. It is interesting to note that for the brightness data of our experiment, $\beta > 1$ for 12 of 15 subjects, and for all 10 subjects who showed significant improvement on the improvement test, a directional result in agreement with all studies

reported using the MR model. A perplexing result taken in conjunction with β magnitudes is that for 14 of our 15 subjects and for all 10 subjects who showed significant improvement on the improvement test, δR was less than unity. In any event, if the bias in our study is due to a relative size factor, the MR model implies brightness assimilation, and the I model implies brightness contrast.

For the interval task, Figure 8 shows a plot of scale values estimated by the interval bias model— $\psi(\delta)$ —as a function of scale values estimated by the no-bias model, $\psi(\delta = 1)$.

The diagonal is the theoretical line for the hypothetical case in which the no-bias model holds exactly. Note that data points for the 8 subjects who failed to show significant improvement through use of a bias parameter lie very close to the diagonal and show no systematic trend. Four of the subjects who showed significant improvement had bias parameter estimates below unity and hence produced data points above the diagonal. This implies that, for these subjects, the middle stimulus of each triple was closer in brightness to the upper standard than was implied by the no-bias model. Three of the subjects who showed significant improvement had bias parameter estimates above unity and produced data points below the diagonal, implying conversely that the middle stimulus was closer in brightness to the lower standard than was implied by the no-bias model. It appears that the effect of the bias parameter on scale values for the interval task is appreciably smaller than the comparable effect for the I model on the ratio task.

Effect of Task on Scale Values

Ekman and Sjöberg (1965) have claimed some support for a log relation between interval and ratio scales, and Stevens and Galanter (1957) have presented evidence that category interval scales are concave down functions of direct ratio scales, but in general the issue of the convergence of scales must be considered an open problem. A key question in the present study is whether the bias models remove some or all of the expected nonlinearity between the interval and ratio scales.

Since a power function with an additive constant can be used to approximate a log function to an arbitrarily close approximation (Fagot, 1966), we make the more general assumption that the interval scale f is a power function of the ratio scale ψ :

$$f = a\psi^m + b. \quad (13)$$

Unfortunately, since the same stimuli were not used for both the interval and ratio task (different stimuli were chosen to generate a wide range of ratios of magnitudes and intervals), Equation 13 cannot be used directly to estimate the exponent m from pairs

Figure 8. Interval bias model scale values as a function of no bias scale values.

(ψ_i, f_i) , and hence we make the classical assumptions that the ratio and interval scales are power functions of physical intensity:

$$\psi = c\Phi^x, \quad (14)$$

$$f = d\Phi^y + e. \quad (15)$$

From Equations 13 and 14, we obtain:

$$f = d'\Phi^{xm} + e'. \quad (16)$$

But, since f is an interval scale, two arbitrary assignments are permitted, and we may rescale by setting $d = d'$ and $e = e'$, in which case we obtain, from Equations 15 and 16, the following interesting relationship among power function exponents:

$$m = y/x. \quad (17)$$

For all subjects combined, a least squares procedure was used for estimating the exponents x and y , and m was computed from Equation 17. Results are shown in Table 3, with the key values of m underlined.

Note that the exponents for both the C and MR models are reasonably close to values typically reported in the direct ratio scaling literature for brightness. The I-model exponent, however, is much lower and close in value to the exponents obtained on the interval task. The values of m indicate that the I-model ratio scale is much more linear with the interval scales than is either the C-model or the MR-model ratio scale. Furthermore, the relation between interval and ratio scales generated by the bias models becomes *more linear* (than the relation

Table 3
Exponents of Power Functions

Interval Task	Exponent "y"
Model 1	.104 (y_1)
Model δ	.080 (y_δ)
Ratio Task	Exponent "x"
C Model	.211 (x_1)
I Model	.133 (x_δ)
MR Model	.253 (x_β)

Values of $m = y/x$			
	x_1	x_δ	x_β
y_1	.494	.786	.411
y_δ	.380	.605	.317

between no-bias scales) if the I model is used to construct the ratio scale, and *less* linear if the MR model is used.

Note that the *differences* between the interval and I-model exponents are small, even though the *ratios* deviate appreciably from unity. However, small (error) differences can substantially affect estimates of low ratios. In any event, the results are sufficiently supporting to suggest a replication using the same stimuli for both interval and ratio tasks, avoiding the confounding influence of two power function assumptions (Equations 14 and 15).

Concluding Remarks

The C model, of course, places the strongest constraints, of the models considered, on the structure of ratio estimation data. The data of the present brightness study, as well as that of earlier studies, indicate that the model should be weakened. Particularly damaging is the systematic tendency for predicted values $\bar{T}_{ij}\bar{T}_{jk}$ to be overestimated relative to observed values $\bar{T}_{ik}$, resulting in rejection of Equation 11 and hence the C model. The same results both as to rejection of Equation 11 and direction of systematic errors were obtained by Eisler (1960) and Goude (1962) for lifted weights, and by Mashhour (1961) for velocity judgments. Fagot and Stewart (1969) also rejected Equation 11 for brightness, using a different judgmental procedure than in the present study, but the inconsistency reported was not of the simple form demonstrated in the above studies. Svenson and Åkesson (1966, 1967) reported data inconsistent with the C model for numerical estimates of line length, circle area, and darkness. Sjöberg (1971) tested the numerical representation for the C model directly, and his data from judgments of lifted weights and areas of circles were generally inconsistent with Equation 5. Thus, there appears to be no basis for retaining the strong assumptions of the C model, even though this theory appears to be accepted by most investigators in the field (Scandinavian psychologists excepted).

The MR model was the first direct weakening of the C model to be investigated. Although the brightness data of the present study raise some questions about the adequacy of the MR model considered in isolation or compared to the I model, there would appear to be sufficient evidence to reject the C model in favor of the MR model. Other research supports this conclusion. Goude (1962), rejecting the C model for judgments of weights and angles, analyzed the data using the MR model, and found good fits in both cases. Sjöberg (1971) used both percent estimation and free ratio estimation (in which two numbers are reported for each pair) with judgments of area of circles and weights. For area of circles, for both types of judgments, results were poor: only 1 of 6 subjects gave satisfactory fit to each of the models. However, using the improvement tests, the MR model demonstrated significant improvement for 3 of the 6 subjects for percent estimation and for 4 of the 6 subjects for free ratio estimation. The results with lifted weights were generally better: for percent estimation, 3 of the 6 subjects showed satisfactory fit for the C model, 5 for the MR model, and 5 subjects showed significant improvement. For free ratio estimation, 4 subjects showed satisfactory fit for the C model, all 6 did for the MR model, and 3 of 6 showed significant improvement. Although Sjöberg did not test Equation 11, his results provided further evidence that the C model needs to be weakened.

A further point concerning the MR model relates to a weakening of that model to produce Sjöberg's (Note 2, 1971) V/S model. As noted previously, the MR model is a special case of the V/S model in which all V/S ratios are assumed constant. Sjöberg (1971) found that for area of circles, using percent estimation and free ratio estimation, 5 of 6 subjects and 2 of 6 subjects, respectively, showed significant improvement on the V/S model relative to the MR model. However, for lifted weights, none of the 6 subjects showed significant improvement for either method of estimation. The difficulty with the V/S model is the very large number of parameters introduced.

No really adequate statistical basis has been provided for comparing the I and MR models on ratio task data. However, goodness-of-fit results do tend to favor the I model, with most subjects showing smaller MSQD and larger reduction in SSD for the I model. This finding, considered in isolation, is not compelling, but takes on added significance in light of other factors favoring the I model. First, the effect on scale values produced by using a bias parameter was much more marked for the I model. This is of considerable importance, since it suggests that scale values estimated by the I model may be relatively invariant under experimental manipulations that affect bias. Second, if we can assume that bias is due to a relative size factor, then the I model implies that

the subject's judgments are affected by brightness contrast, whereas the MR model implies brightness assimilation. This would appear to lend support to the I model, since perceptual theory and experiment would lead us to expect a brightness contrast effect. Third, the relation between interval and ratio scales shows the least curvilinearity if the I model is used to construct the ratio scale, and the most curvilinearity if the MR model is used to construct the ratio scale (i.e., even more curvilinearity than if the C model is used).

With respect to the interval task, less than half of the subjects showed significant improvement through use of the bias parameter. Of the 7 who did, 4 had bias parameter estimates below unity and 3 above unity, suggesting that for judgments of triples of stimuli, unlike pairs, a single perceptual process such as brightness contrast or assimilation, did not operate for all subjects. The test of the interval product condition (Equation 12, Figure 6) contrasted with the test of the ratio product condition (Equation 11, Figure 5) lends support to the proposition that judgments of triples of (brightness) stimuli are less influenced by response bias than are judgments of pairs. The interval estimation data are viewed as generally supportive of recent efforts directed toward the development of direct interval scaling theory.

REFERENCE NOTES

1. Mashhour, M. On the validity of scales derived by ratio and magnitude estimation methods. Report of the Psychology Laboratory, University of Stockholm, 1961, No. 105.
2. Sjöberg, L. On ratio estimation. Report of the Psychology Laboratory, University of Stockholm, 1965, No. 191.
3. Svenson, O., & Åkesson, C. A. Fractional and multiple estimates in ratio scaling. Report of the Psychology Laboratory, University of Stockholm, 1966, No. 202.
4. Svenson, O., & Åkesson, C. A. A further note on fractional and multiple estimates in ratio scaling. Report of the Psychology Laboratory, University of Stockholm, 1967, No. 224.

REFERENCES

- DRAPER, N. R., & SMITH, H. *Applied regression analysis*. New York: Wiley, 1966.
- EISLER, H. Similarity in the continuum of heaviness with some methodological and theoretical considerations. *Scandinavian Journal of Psychology*, 1960, 1, 69-81.

- EKMAN, G., & SjöBERG, L. Scaling. *Annual Review of Psychology*, 1965, 16, 451-474.
- ESKILDSEN, P. An apparatus for studies of brightness which controls variation of intensity with time. *American Journal of Psychology*, 1963, 76, 321-323.
- FAGOT, R. F. Alternative power laws for ratio scaling. *Psychometrika*, 1966, 31, 201-214.
- FAGOT, R. F., & STEWART, M. R. Tests of product and additive scaling axioms. *Perception & Psychophysics*, 1969, 5, 117-123.
- FAGOT, R. F., & STEWART, M. R. Test of a response bias model of bisection. *Perception & Psychophysics*, 1970, 7, 257-262.
- GOUDE, G. *On fundamental measurement in psychology*. Stockholm: Almqvist & Wiksell, 1962.
- HARTLEY, H. O. The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares. *Technometrics*, 1961, 3, 269-280.
- HEMMERLE, W. J. *Statistical computations on a digital computer*. Massachusetts: Blaisdell, 1967.
- KRANTZ, D. H. A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, 1972, 9, 168-199.
- MARLEY, A. A. J. Internal state models for magnitude estimation and related experiments. *Journal of Mathematical Psychology*, 1972, 9, 306-319.
- PFANZAGL, J. In cooperation with V. Baumann and H. Huber. *Theory of measurement*. New York: Wiley, 1968.
- SjöBERG, L. Three models for the analysis of subjective ratios. *Scandinavian Journal of Psychology*, 1971, 12, 217-240.
- STEVENS, S. S. On the psychophysical law. *Psychological Review*, 1957, 64, 153-181.
- STEVENS, S. S., & GALANTER, E. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, 54, 377-411.
- SUPPES, P., & ZINNES, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963. Chap. 1.

NOTE

1. There are other ad hoc ways in which a bias parameter could be introduced into the numerical representation for interval judgments, but Equation 3 has the advantage of being a theoretical consequence of the generalization of the bisection representation (Equation 2).

(Received for publication September 19, 1974;
revision received November 25, 1974.)