

Bias effects on magnitude and ratio estimation power function exponents

ROBERT F. FAGOT and ROBERT POKORNY
University of Oregon, Eugene, Oregon

A bias model of relative judgment was used to derive a ratio estimation (RE) power function, and its effectiveness in providing estimates of exponents free of the effects of standards was evaluated. The RE bias model was compared with the simple RE power function that ignores bias. Results showed that when bias was not taken into account, estimates of exponents exhibited the usual effects of standards observed in previous research. However, the introduction of bias parameters into the RE power function virtually eliminated these effects. Exponents calculated from "equal-range segments" (e.g., low stimulus range vs. high stimulus range) judged by magnitude estimation (ME) were examined: the effects of equal-range segments on exponents were much stronger for ME than standards were for RE, using the bias model.

The power function (S. S. Stevens, 1957), based on ratio scaling methods, appears to be firmly established as a psychophysical law. Yet context effects on the exponent of the power function are ubiquitous—depending, for example, on scaling method, range and spacing of stimuli, standard and modulus, and sequence of stimuli (Poulton, 1968, 1979). In addition, there are numerous reports of large individual differences in exponents for the same modality (e.g., Marks & Cain, 1972; Marks & J. C. Stevens, 1966; J. C. Stevens & Guirao, 1964).

S. S. Stevens's concern with the effect of the standard on the exponent led to his adoption of the "no-standard" procedure (see S. S. Stevens, 1956, and 1975, p. 27). Marks (1974b) also sees little reason to use a standard in magnitude estimation experiments, and Cross (1982) views "current practice" among psychophysicists as imposing minimal constraints on the scaling task. This position would be more convincing if magnitude estimation without a standard were free of context effects, but such is not the case. For example, range and sequence effects are present for magnitude estimation with or without a standard (Poulton, 1968, 1979). R. Teghtsoonian and M. Teghtsoonian (1978) reported that, for both loudness and distance judgments, intramodal range exerts a systematic effect on the exponent, which results from magnitude estimation without a standard; and Cross (1973) reported sequential dependencies in magnitude estimation of loudness, without a standard or modulus.

It is argued by many investigators, including the present authors, that an adequate theory of psychophysics must

develop models that account for, rather than ignore or average out, context effects. One example is the work of Parducci (1982), who proposes a range-frequency theory that describes well the effects of stimulus range, spacing, and frequency for category ratings. R. Teghtsoonian (1973) and R. Teghtsoonian and M. Teghtsoonian (1978) propose a model utilizing stimulus and response ratios that gives a good description of the effects of intramodal stimulus-range variation, making it possible in theory to isolate the exponent as an invariant, relative to the effects of range variation.

It is in this spirit that we propose in the present study to use a model of relative judgment (Fagot, 1981), with which we will attempt to account for the biasing effects of standards, and from which a power function will be derived. To the extent that the theory is correct, the exponents of the derived power function should be relatively free of the effects of standards.

For magnitude estimation (no standard) we shall examine the effects of the disjoint "equal-range segment" used to construct the power function—for example, the lower half versus the upper half of the stimulus range. Many studies of ratio estimation have reported on the effects of standards on the exponent of the power function. The purpose in introducing the equal-range condition is to ask a similar question of magnitude estimation: does the estimation of the exponent that results from magnitude estimation depend on the location of the segment? Since the segments are equal in range, the segment effect will not be confounded with the well-studied stimulus range effect.

As a preliminary step, the fit of individual power functions will be assessed. Bias effects on exponents for each method will be evaluated in terms of statistical tests, the effect size (including that of the "subject effect"), and the stability of exponents across conditions. The conditions for ratio estimation will be *response type* (fractional or multiple judgments) and *standards*. The conditions for magnitude estimation will be *equal-range segments*.

This research was supported in part by Public Health Service Grant 5R01 MH 20449-06. This paper has benefited considerably from critical suggestions by M. Teghtsoonian, as Action Editor, and two anonymous reviewers. Robert Pokorny is now at Brooks Air Force Base, AFHRL/MOD, San Antonio, TX 78235-2601. Requests for reprints should be sent to Robert Fagot, Department of Psychology, University of Oregon, Eugene, OR 97403-1227.

THEORY

By magnitude estimation (ME), we mean a procedure in which the stimuli are presented one at a time, and the overt response is the subjective magnitude M_a of some stimulus a . For ratio estimation (RE), the stimuli are presented two at a time, and the overt response R_{ab} is the "ratio" of the magnitude of stimulus a (the comparison stimulus) to that of stimulus b (the standard).

The well-known power law states that sensation magnitude (ψ) grows as a power function of physical magnitude (S); i.e. $\psi = cS^k$. Since ψ is unobservable, a transformation (the judgment function, or the scaling model) is required that relates ψ to the overt response, from which one may derive a power function expressed in terms of overt response and physical magnitude S .

For magnitude estimation, there does not exist a scaling theory that identifies the required transformation, except for the simplistic notion that response (M) is proportional to ψ (see Shepard, 1981). Hence, one either makes this assumption or assumes outright that the overt response is a power function of S ; thus,

$$M = cS^m, \tag{1}$$

where M is taken as some average, usually the geometric mean, of several overt responses of the magnitude of a specific stimulus. That is the practice to be followed in this paper, and Equation 1 is the form of the power function that will be tested for magnitude estimation (an additive constant will not be needed because stimuli will not be near threshold).

In the case of ratio estimation, it is usually assumed that the overt response is a monotone function of sensory ratios, resulting in the judgment function $R_{ab} = g(\psi_a/\psi_b)$. If it is assumed that the subject is literally forming subjective ratios in accordance with instructions, then g becomes the identity transformation, and we get

$$R_{ab} = \psi_a/\psi_b, \tag{2}$$

called the classical (C) model by Fagot (1978). From $\psi = cS^r$ and (2), the C-model power function is

$$R_{ab} = (S_a/S_b)^r \tag{3}$$

(Baird & Noma, 1978; Fagot, 1982). Note that the multiplicative constant, c , cancels, and Equation 3 contains only one unknown, the exponent r . We shall refer to Equation 3 as the RE(C) power function.

Fagot (1981) proposed the following judgment function in an attempt to deal with the biasing effects of standards:

$$R_{ab} = \alpha_{bH}(\psi_a/\psi_b) \tag{4}$$

$(a < b)$

$$R_{ba} = \alpha_{aL}(\psi_b/\psi_a)$$

where the first subscript on R denotes the comparison stimulus and the second the standard. α_{bH} is the bias for standard b when the standard is larger than the compari-

son stimulus, and α_{aL} is the bias for standard a when the standard is smaller than the comparison stimulus. Equation 4 is called the relative bias/directional standard (RBDS) model. The RBDS model allows each standard to have two biasing effects, depending on the location of the standard relative to the comparison stimulus (see Fagot, 1981, p. 183, for details). Considerable evidence (Fagot, 1981) indicates that the RBDS model provides a good account of the biasing effects of standards.

The RBDS power function is then:

$$R_{ab} = \alpha_{bH}(S_a/S_b)^r \tag{5}$$

$(a < b)$

$$R_{ba} = \alpha_{aL}(S_b/S_a)^r$$

Equation 5 requires estimates of the bias parameters—either prior estimates from the scaling experiments, or the simultaneous estimation of the bias parameters and the exponent in the construction of the power function. But it is possible to use the axiom of the RBDS model to derive an RBDS power function in which the bias parameters do not appear, as shown below.

The axiom for the ratio estimation RBDS model is the following *ratio of ratios* condition (Fagot, 1981):

$$\frac{R_{aw_1}}{R_{bw_1}} = \frac{R_{aw_2}}{R_{bw_2}} = \dots = \frac{R_{aw_n}}{R_{bw_n}}$$

$(a, b < w_i \text{ or } a, b > w_i, \text{ for all } i = 1, 2, \dots, n).$ (6)

The ratio of ratios condition constrains the stimuli so that the two comparison stimuli (a, b) must both be either below or above each standard w_i (i.e., each ratio contains two fractional or two multiple judgments).

Equation 5, together with the ratio of ratios condition, entails the following simple expression for the RBDS power function for comparison stimuli a and b , and standards w_i :

$$(R_{aw_i}/R_{bw_i}) = (S_a/S_b)^r$$

$$(a, b < w_i \text{ or } a, b > w_i, \text{ for all } i = 1, 2, \dots, n). \tag{7}$$

Equation 7 will be referred to simply as the RE power function, and Equation 5 as the RE(BP) power function. All three RE power functions (Equations 3, 5, and 7) will be evaluated.

Because of the constraints placed on the comparison stimuli relative to the standards, the bias parameters in Equation 5 cancel in Equation 7. The RBDS model, through the constraints placed on Equation 7, attempts to "remove" the effect of response bias on the exponent of the power function. Equation 7 implies that the exponent is invariant, relative to the effects of standards, based on the ratio of ratios condition.¹

An alternative ME power function, derived from Equation 1, is:

$$(M_a/M_b) = (S_a/S_b)^m, \tag{8}$$

where the ratios M_a/M_b are formed from two independent

magnitude estimates. The multiplicative constant C (scale factor) in Equation 1 vanishes in Equation 8, resulting in an ME power function with only the exponent to estimate, as in the case of RE. Equation 8 will be useful in comparing ME and RE.

METHOD

Twelve subjects performed a loudness and a heaviness estimation task, completing them in two 1.75-h sessions, 1 week apart. Six of the subjects judged loudness first and heaviness second; the order was reversed for the other 6.

The loudness and heaviness sessions were similar. Both sessions began with magnitude estimation, using nine stimulus levels. The equal-range segments used in data analysis were defined by the experimenter but not presented as such to the subject. All nine stimuli were presented in six different random orders, for each of the modalities. No standard or modulus was assigned.

The ratio estimation procedure used the third, fifth, and seventh stimulus levels as standards, and trials were blocked by standard. The order of standard for the ratio estimation trials was either 5, 7, 3, for 6 subjects, or 5, 3, 7, for the other 6; the same order of standards was used for the two modalities. The order of modality was counterbalanced within each order of standards. All nine stimuli were presented, relative to each standard, in six different random orders.

For both procedures, each estimate was made six times, and the initial estimate for each stimulus was treated as a practice trial.

Subjects

The subjects were students from the University of Oregon and nonstudents from the community. They were all right-handed and had no prior experience making psychophysical judgments. They were paid \$3 per hour.

Apparatus and Stimuli

Heaviness. A classroom chair with an extended right arm and an attached plywood board formed a 50×60 cm desktop, in which two holes were drilled 13 cm and 22 cm away from the right edge, and 1.5 cm from the far edge (relative to the subject). Through each hole, a nylon fishing line connected an inverted plastic cup above to weights below the desktop. The cup was 2 cm high, 1.5 cm wide at its base, and 1 cm wide at its top. The weights were plastic containers filled to the correct weight with lead shot and attached to the fishing line with hooks. The containers were concealed from the subjects' view. The experimenter sat on the floor in front of the subject in order to hang the appropriate weights on the fishing lines. Nine weights, ranging from 75 to 500 g in equal logarithmic steps, were used.

Loudness. The subjects listened to auditory signals presented dichotically through Koss 70 headphones and generated by a Grason-Stadler 950-C signal oscillator (1000-Hz tone). Stimulus duration was controlled by Hunter timers. The experimenter sat in a room adjacent to the subject's room, with the door slightly ajar to allow for easy communication. The experimenter initiated signal presentation by pressing a start key to activate the timers, and monitored the output of the signal generator with a Heathkit multimeter. The nine signals were separated by 6-dB steps. The least intense sound was set to be 30 dB above the average absolute threshold of 4 pilot subjects. Thus, the signals ranged from 30 to 78 dB above threshold. All signal durations were 1 sec; the interstimulus interval between the ratio estimation signals was 1 sec.

Procedure

Heaviness. The subjects began the heaviness session with free-modulus magnitude estimation. The subjects assigned a number proportional to the perceived heaviness of each weight. They were

instructed to grasp the plastic cup with the thumb, pointer and index fingers, and to pull it straight up, keeping the elbow on the desk while lifting the entire forearm. They were allowed to lift the weight as many times as necessary before verbally reporting the weight's heaviness. The nine weights were presented in six different random orders. The first estimate of each weight was considered a practice trial.

The subjects then participated in ratio estimation. A standard weight, with an assigned value of 100, and a comparison weight were presented on each trial. The subjects reported the heaviness of the comparison weight relative to that of the standard, given the usual instructions. For example, they were told that "if the stimulus feels twice as heavy, report 200, and if it feels one-fourth as heavy, report 25." They were instructed to grasp the weight as described for magnitude estimation. The standard weight had to be lifted first, and the comparison had to be the last weight lifted before a response; otherwise, the subjects could lift the weights in any sequence. Each standard was paired with the other eight stimuli in six different random orders for a total of 48 estimates; the first estimate of each standard-comparison pairing was considered practice.

Judged loudness. The loudness judgment experimental session was identical to the heaviness session in (1) levels of stimuli, and (2) order of methods.

On magnitude estimation trials, the experimenter warned the subject of signal onset about 1 sec before the signal was presented. The subject either reported a number that was proportional to the loudness of the sound or asked for the sound to be repeated. As in heaviness estimation, nine stimulus levels were presented in six different random orders; the initial estimate of each stimulus level was considered practice. No standard or modulus was assigned.

For ratio estimation, the standard signal, called 100, was presented first, followed after a 1-sec interval by the comparison signal. The subjects were instructed to report the loudness of the second signal. Instructions were the same as those for heaviness. If the subject requested, the sequence of two tones was repeated. Each standard was paired with the other eight stimuli in six different random orders for a total of 48 estimates with each standard; the first comparison-standard pairing was considered practice.

Regression Analysis

For each of the power functions, ordinary least squares (OLS) procedures provided estimates of the exponents and tests of deviations from individual power functions.

For ratio estimation, there were nine comparison stimuli and three standards (S_3 , S_5 , S_7) selected from the comparison stimuli, for a total of 24 R_{ab} and R_{ba} estimates, with five replications of each. The geometric mean of the five observations was used as the estimate of R_{ab} .

The log-log form of Equation 3 provided estimates of exponents and tests of the RE(C) power function. The error term for the F test was the within-mean square, based on five estimates of each of the 24 R_{ab} estimates.

The ratio of ratios condition (Equation 6) incorporated into the RE power function (Equation 7) provided a means of testing relative to scale precision, or more specifically, relative to the necessary condition of the RBDS model. Equation 7 entails that for each fixed S_i/S_j , there will be two or three equal R_{aw}/R_{bw} . For example, (1) $\ln(R_{13}/R_{23}) = \ln(R_{15}/R_{25}) = \ln(R_{17}/R_{27}) = r \ln(S_1/S_2)$, and (2) $\ln(R_{15}/R_{35}) = \ln(R_{17}/R_{37}) = r \ln(S_1/S_3)$, and expressed in compact form as the regression of Y on X :

$$(1) Y_{12i} = r X_{12} \quad (i = 3, 5, 7)$$

$$(2) Y_{13i} = r X_{13} \quad (i = 5, 7)$$

The three repeated observations of Y in (1) and the two repeated observations of Y in (2) contribute to the error mean square (*within-axiom* variance) for testing lack of fit to the RE power function.

Hence, the lack of fit will measure deviations from the power function relative to deviations from the scaling function. Thus, a scale with high precision (small within-axiom variance) will provide a more powerful test of lack of fit. The constraints of Equation 7 reduced the total set of 24 (a,b) pairs to 15. Thirteen of the (a,b) pairs had two repeated observations, and two of the pairs had three.

For magnitude estimation, each subject made five independent estimates M_a of each of nine stimuli. From Equation 1, the regression of $\ln M$ on $\ln S$, by OLS, provided a test of the fit of the ME power function. The error term for the F test is the within-mean square based on five estimates of each of the nine stimuli (separately for each subject).

The final regression analysis is based on Equation 8, wherein the regression of $\ln(M_a/M_b)$ on $\ln(S_a/S_b)$, via OLS, provided an estimate of the exponent m and a measure of fit R^2 . The utility of Equation 8 is its use of the same regressor variable as in Equation 7 for RE, [with the (a,b) pairs determined by the constraints imposed by Equation 7], essential for the comparability of exponent and R^2 values. Equation 8 cannot be used to test for lack of fit to the ME power function since there is no mean square within. Hence the test of lack of fit to the ME power function will be based on Equation 1, and the estimated exponents and R^2 measures on Equation 8. R^2 is the proportion of variation in log response accounted for by linear regression (i.e., the power function). Specifically, $R^2 = SS(\text{reg})/[SS(\text{reg}) + SS(\text{dev})]$.

RESULTS

Results are organized under two comparisons: (1) the RE power function (Equation 7) with the ME power function (Equations 1 and 8); and (2) the RE power function with and without bias parameters (Equations 5 and 3, respectively). Analysis is made of (1) exponent estimates and goodness-of-fit tests, (2) bias and subject effects, and (3) stability of exponents. Separate regression analyses were carried out for each of the subsets of stimuli defining the conditions for RE (standards, response types) and ME (equal-range segments).

Regression analyses were carried out for each power function (Equations 1, 3, 5, 7, and 8), each modality (loudness, heaviness), and each of the 12 subjects.

The RE Power Function (Equation 4) Compared with the ME Power Function (Equations 1 and 8)

Exponent estimates and goodness-of-fit tests. Table 1 shows estimates of exponents, R^2 , and the results of F tests of lack of fit to each power function. The last two rows show mean exponents for each method and modality, as well as the standard deviation of each set of individual exponents. RE and ME appear to provide good fits to a power function. For RE, only 1 subject showed significant lack of fit for heaviness, and 3 for loudness. For ME, 2 subjects showed significant lack of fit for heaviness, and 3 for loudness.

The loudness exponents in Table 1 are all below 1, as expected, and the means for ME and RE are quite close. However, the exponents are lower than the 0.6 values commonly found for sound pressure. The exponents for heaviness differ in two important respects. First, the mean exponent for RE (1.35) is significantly higher than for ME (1.02): [$t(11) = 3.07, p < .025$]. Ten of 12 subjects had higher exponents for RE. Second, all but one of the subjects for RE had an exponent greater than 1, whereas for ME only 5 of the 12 subjects had exponents above 1.

R^2 values indicate that the RE power functions are better fits than those for ME. For heaviness, mean R^2 is appreciably higher for RE (.94 vs. .83), and 11 of 12 subjects showed higher R^2 for RE. There is not as much difference in R^2 values between RE and ME for loudness, although the mean R^2 is higher for RE, and 9 of 12 subjects show higher R^2 for RE.

Bias and subject effects. The effect of bias and subject on exponents for both RE and ME are evaluated in terms of statistical significance, effect size (ES) as measured by ω^2 , and stability of exponents across conditions. The conditions for RE are (A1) response types: fractional judgments (comparison stimulus below standard) and multiple judgments (comparison stimulus above standard), and (A2) standards. For ME two disjoint equal-range segment

Table 1
Exponent Estimates and Tests of Fit of Power Functions

Subject	Loudness						Heaviness					
	RE			ME			RE			ME		
	$F\ddagger$	r	R^2_R	$F\§$	m	R^2_M	$F\ddagger$	r	R^2_R	$F\§$	m	R^2_M
D.A.	1.53	.54	.93	3.81†	.51	.83	0.16	1.19	.99	1.54	.63	.72
C.F.	1.82	.36	.90	1.81	.37	.88	1.03	1.08	.93	0.92	.68	.95
C.L.	2.66*	.26	.86	4.12†	.38	.84	0.89	1.71	.85	1.08	.80	.80
C.W.	1.67	.63	.95	1.09	.50	.96	0.80	2.41	.95	1.57	1.75	.84
E.E.	0.98	.46	.96	2.43*	.38	.83	1.02	0.87	.94	1.11	.55	.87
J.F.	1.95	.40	.94	1.25	.48	.94	0.91	1.54	.99	1.78	.89	.84
J.M.	2.44*	.50	.86	0.48	.54	.91	0.73	1.42	.94	0.62	1.08	.88
M.N.	0.40	.37	.98	0.36	.42	.97	0.69	1.11	.96	2.34*	1.63	.80
M.S.	0.51	.39	.96	0.86	.39	.94	1.20	1.48	.94	1.41	1.23	.88
N.N.	0.67	.25	.92	0.50	.34	.92	1.09	1.01	.88	2.83*	.89	.76
R.R.	2.97*	.38	.89	1.82	.52	.79	4.69†	1.37	.94	0.62	1.45	.77
T.R.	0.91	.33	.95	1.93	.26	.85	2.22	1.03	.93	0.37	.62	.85
<i>M</i>		.41	.92		.42	.89		1.35	.94		1.02	.83
<i>SD</i>		.11			.08			.42			.41	

Note— R^2_R is based on Equation 7, and R^2_M is based on Equation 8. RE = ratio estimation, ME = magnitude estimation.

* $p < .05$ † $df = 14,17$ (Regression model: $Y = \beta X$).
 ‡ $p < .01$ § $df = 7,36$ (Regression model: $Y = \beta X + \alpha$)

conditions were defined²: (B1) low (Stimuli 1-4) and high (Stimuli 6-9); and (B2) low (1-3), medium (4-6), and high (7-9).

All exponents were estimated by OLS separately for each condition, using the log-log form of Equation 7 for RE and the log-log form of Equation 8 for ME. Note that Equation 8 uses the same regressor variable as Equation 7 ($\ln S_a/S_b$). Thus, while for ME there are no true fractional or multiple judgments (since stimuli are judged one at a time), we define comparable response types, namely fractional ratios M_a/M_b , and multiple ratios M_b/M_a ($a < b$). This means that response type and equal-range segment condition B1 are based on the same six (a, b) pairs. In addition, by using only Standard 5 to estimate exponents for response type, the number of observations for each data point was equalized for RE and ME. Hence on the basis of Equation 7 for RE and Equation 8 for ME, the estimation of exponents for response type and equal-segment condition B1 was strictly comparable, since six data points with five observations for each data point were used in both cases, with the same levels on the regressor variable. In addition, the dependent variable had the same form for both methods, namely a ratio of responses. Although standards (A2) and equal-range segment condition B2 both consist of three levels, A2 and B2 are not directly comparable since the exponents for A2 are based on eight data points, whereas the exponents for B2 are based on only three data points.

All subject and bias effects were evaluated by means of subject \times condition repeated-measures designs. The effect size was estimated by ω^2 for one-factor repeated-measures designs assuming an additive model (cf. Dodd & Schultz, 1973; Susskind & Howland, 1980).

Table 2 (A1) shows that for RE, exponents were higher for multiple than fractional judgments for both loudness and heaviness, a difference consistent with research reported elsewhere (Poulton, 1968). However, not all subjects showed this trend (7 of 12 for loudness, 9 of 12 for heaviness), and the difference was not significant for either loudness or heaviness. The effect of standards (A2) also was not significant for either loudness or heaviness, although previous reports (Beck & Shaw, 1961, 1965; Hellman & Zwislocki, 1961; S. S. Stevens, 1956; J. C. Stevens & Tulving, 1957) indicate that the exponents tend to be higher for the middle than for the two extreme standards. In sum, no bias effects were significant for RE. For ME, the effect of equal-range segments was significant for heaviness (B1) and for loudness (B2), but not in the other two cases. Ten of 12 subjects showed the mean trend for loudness in condition B2 [exp(H) highest].

Two estimates of ES are given in Table 2, for both bias condition (ES_c) and subjects (ES_s)³. The first entry in the appropriate cell of the table gives ES as a proportion of total variance, and the entry in parentheses following gives partial ES . The most striking results in Table 2 are the estimates of ES as a proportion of total variability. The estimates of ES_c are near zero even in those cases of statistical significance, except in the case of ME of loudness (B2). On the other hand, the estimates of ES for subjects is quite high for RE, but appreciably lower for ME. The results are clearer for RE: subjects account for an average of about two thirds of the variability in exponents, but bias condition for practically none (mean = .03). The contrast between subject and bias effect is not so strong for ME, although for three of the four cases the ES_c is, as for RE, very low (mean = .11), with subjects account-

Table 2
Effects of Bias on Ratio Estimation and Magnitude Estimation Exponents

A. Ratio Estimation			B. Magnitude Estimation		
	LOUD	HEAV		LOUD	HEAV
A1. Response Type: Fractional (F) and Multiple (M)			B1. Equal-Range Segment: Low (1-4) and High (6-9)		
exp(F)	0.43	1.25	exp(L)	0.40	1.15
exp(M)	0.46	1.49	exp(H)	0.47	0.83
$F_c(1,11)$	0.40	1.96	$F_c(1,11)$	1.75	5.36*
$F_s(11,11)$	3.30*	2.56	$F_s(11,11)$	0.78	3.08*
ES_c	0.00 (0.00)	0.02 (0.04)	ES_c	0.03 (0.03)	0.08 (0.15)
ES_s	0.54 (0.54)	0.43 (0.44)	ES_s	0.00 (0.00)	0.47 (0.51)
A2. Standards (3, 5, 7)			B2. Equal-Range Segment: Low (1-3), Medium (4-6), High (7-9)		
exp(3)	0.37	1.44	exp(L)	0.38	0.98
exp(5)	0.44	1.37	exp(M)	0.35	1.20
exp(7)	0.39	1.23	exp(H)	0.57	0.82
$F_c(2,22)$	3.45‡	3.08	$F_c(2,22)$	12.01†	1.73
$F_s(11,22)$	7.45†	11.64†	$F_s(11,22)$	2.70*	2.26*
ES_c	0.04 (0.12)	0.02 (0.10)	ES_c	0.28 (0.38)	0.03 (0.04)
ES_s	0.65 (0.68)	0.76 (0.78)	ES_s	0.26 (0.36)	0.29 (0.29)

Note—exp = mean exponent; ES = effect size estimated by ω^2 . Subscripts on F and ES are: c = bias condition, S = subjects. Figures in parentheses after ES entries are partial ES .⁴ †Degrees of freedom adjusted to 1 and 14 due to violations of equal covariance assumption: $p > .05$ (see Myers, 1979, pp. 173-174). * $p < .05$. ‡ $p < .001$.

ing for an average of 38% of the variability in exponents for heaviness, but only 13% for loudness.⁴

Stability of exponents. Both the Pearson *r* and the coefficient of identity *e'* are used in our analysis to evaluate the stability of the exponent.⁵ Although *r* is not sensitive to linear context bias (e.g., an increase in exponent due to standard that affects all subjects equally), its statistical significance can be assessed and its wide use allows comparison with other studies. The coefficient *e'* (proposed by Zegers, 1986; based on Zegers and ten Berge, 1985) assesses the extent to which individual exponents are the same in both conditions; *e'* is sensitive to any change induced by bias and is always less than or equal to *r*. However, no statistical test is available for *e'*. Table 3 presents *e'* and *r* for within modalities and *r* for between modalities.

Table 3 shows that the stability of exponents for bias condition is much greater for RE than for ME, except for heaviness in Conditions C1 and D1, where the correlations are about equal. The stability of exponents across standards (C2) is uniformly high. The correlations for loudness in Conditions C1 and D1 are particularly damaging for ME, but the correlations for Segment Condition D2, although very low, should be interpreted cautiously due to the small number of data points. Overall, all correlation coefficients (*r*) were significant for RE, but only three of eight were for ME. In general, the correlations for ME are surprisingly low.

It has been pointed out that, for our design, ME always preceded RE. The justification for omitting the reverse

order must be viewed in the context of recent work on the formulation of the psychophysical function in terms of stimulus and response ratios. R. Teghtsoonian (1973) and R. Teghtsoonian and M. Teghtsoonian (1978), for example, use just such a modification of S. S. Stevens' law to describe the intramodal range effect. Krantz (1972) presented two kinds of ratio scaling theory: *mapping theory*, based on the traditional view that subjects' judgments are mediated by sensation, and *relation theory*, based on a proposal by Shepard (1981) that subjects' judgments are mediated by perceived relations relative to pairs of stimuli. Mapping theory implies that stimuli are mapped into the subjective domain one at a time, whereas relation theory proposes that stimuli are mapped into the subjective domain in pairs, evoking a sensation ratio. Contrasting the two approaches requires that a clear distinction be made to the subject between single stimulus presentation (ME) and pair presentation (RE). But preceding ME with RE may have encouraged the subjects to judge a stimulus relative to the preceding stimulus, and possibly resulted in a mapping of stimuli in pairs, rather than one at a time, for ME.

Nevertheless, it might be argued that the less favorable results for ME may have been due to less familiarity with the stimulus distribution. We have examined three lines of indirect evidence that discount the order effect as a major contributor to the poorer results for ME.

If amount of experience with the distribution of stimuli was a major factor in the ME results, then this should be demonstrated in a comparison of early and late obser-

Table 3
Stability of Exponents

A. Between Methods (RE and ME)			B. Between Modalities (Loudness and Heaviness)		
	LOUD	HEAV			
<i>r</i>	.66†	.59*	RE: <i>r</i>	= .47	
<i>e'</i>	.63	.44	ME: <i>r</i>	= .48	
C. Ratio Estimation			D. Magnitude Estimation		
C1. Fractions and Multiples			D1. Equal-Range Segment: Low (1-4), High (6-9)		
	LOUD	HEAV		LOUD	HEAV
<i>r</i>	.54*	.54*	<i>r</i>	-.15	.57*
<i>e'</i>	.51	.40	<i>e'</i>	-.10	.41
C2. Standards (3, 5, 7)			D2. Equal-Range Segment: Low (1-3), Medium (4-6), High (7-9)		
	LOUD	HEAV		LOUD	HEAV
(3,5):			(L,M):		
<i>r</i>	.62*	.69†	<i>r</i>	.20	-.03
<i>e'</i>	.51	.68	<i>e'</i>	.17	-.03
(3,7):			(L,H):		
<i>r</i>	.54*	.79†	<i>r</i>	.34	.15
<i>e'</i>	.53	.70	<i>e'</i>	.14	.14
(5,7):			(M,H):		
<i>r</i>	.88‡	.89‡	<i>r</i>	.58*	.60*
<i>e'</i>	.79	.83	<i>e'</i>	.28	.47

Note—*r* and *e'* based on N=12 subjects. *e'* = coefficient of identity.⁵ **p* < .05. †*p* < .01. ‡*p* < .001. (*df* = 10, one-tail *t* test)

vations. To test this, the first two observations (T_1) were compared with the last two (T_2). Three comparisons were made between T_1 and T_2 . If experience was a major factor, then, first, the correlation between low and high segments (D1, Table 3) ought to have been higher for T_2 (this result was not found); second, reliability (correlation between T_1 and T_2) ought to have been relatively low; and third, error sum of squares (SS_w) ought to have been higher for T_1 . In fact reliability coefficients were relatively high, and SS_w was not higher for T_1 .

We interpret these results as showing that the dominant influence operating for ME is the systematic influence of equal-range segments, not pure "unreliability," and further, that the indirect evidence does not support the hypothesis that less familiarity with the stimulus distribution was responsible for the poorer results for ME.

Comparison of the RE(C) and RE(BP) Models

If the favorable results for RE in Tables 1-3 are to be attributed specifically to the bias form of the RE model, then the case would be strengthened if the bias model can be shown to be superior to the no-bias RE model—that is, the RE(C) model (Equation 3). This analysis follows.

All exponents for the RE(C) model were estimated by OLS, separately for each condition, using the log-log form of Equation 3. To assess the success of the bias model in reducing the effect of standards and response type on exponents, the bias parameters were introduced explicitly into the power function by using the log-log form of Equation 5, separately for each condition. The utility of Equation 5 is its use of the same dependent variable ($\ln R_{ab}$) and regressor variable ($\ln S_a/S_b$) as that used in the RE(C)

model (Equation 3). This provides a direct comparison of the bias and no-bias models using the same (X, Y) values in the regression equation.

The bias parameters (α_x) were estimated from the formula $\alpha_x = R_{ax}R_{xb}/R_{ab}$ (see Fagot, 1981). There are six bias parameters ($\alpha_{3L}, \alpha_{3H}, \alpha_{5L}, \alpha_{5H}, \alpha_{7L}, \alpha_{7H}$), but note that the bias parameters are estimated directly from the observed R_{ab} prior to the construction of the power function. Thus the bias parameters are not estimated jointly with the exponent via the regression equation with the aim of minimizing errors in the fit to the power function, but rather they appear as *known constants* in the regression of $\ln R_{ab}$ on $\ln(S_a/S_b)$.

Goodness-of-fit tests for RE(C) model. *F* tests of goodness-of-fit to the RE(C) model showed that *all* subjects for both loudness and heaviness had significant lack of fit to Equation 3. Results were particularly poor for loudness: for all 12 subjects, the *F* test was significant at the .001 level. For heaviness, the results were not much better, with 8 of the 12 subjects showing significant lack of fit at the .001 level, 1 at the .01 level, and 3 at the .05 level. These results for the RE(C) model are in marked contrast to the favorable results for both RE and ME (Table 1), and supports the utility of the RE bias model.

Mean exponents, effect size, and stability of exponents. The results of the comparison between the RE(C) and RE(BP) models with respect to mean exponents, effect size, and stability of exponents, are presented in Table 4.

Table 4B shows that all bias effects were highly significant for the RE(C) model, and none of the bias effects was significant for the RE(BP) model. As expected,

Table 4
Comparison of RE(C) and RE(BP) Models: Mean Exponents, Effect Size, and Stability of Exponents

		Loudness		Heaviness	
		RE(C)	RE(BP)	RE(C)	RE(BP)
A. Mean Exponents					
Response	F	.29	.38	1.49	1.46
Type:	M	.41	.35	1.21	1.47
Standards:	3	.26	.35	1.41	1.45
	5	.42	.38	1.44	1.47
	7	.39	.37	1.24	1.48
B. Effect Size					
Response	F_c	16.77†	1.43	17.67†	0.10
Type:	F_s	3.07*	4.41*	14.30†	38.17†
	ES_c	.24 (.40)	.01 (.02)	.08 (.41)	.00 (0)
	ES_s	.38 (.51)	.53 (.63)	.80 (.87)	.93 (.96)
Standards:	F_c	29.84†	0.73	5.38*	0.33
	F_s	9.96†	11.90†	25.25†	75.08†
	ES_c	.29 (.62)	.00 (0)	.03 (.20)	.00 (0)
	ES_s	.53 (.75)	.79 (.79)	.87 (.89)	.96 (.96)
C. Stability of Exponents (e')					
	F,M	.30	.61	.72	.95
	3,5	.30	.71	.95	.98
	3,7	.31	.77	.82	.94
	5,7	.79	.82	.75	.96

* $p < .025$. † $p < .001$.

all subject effects were significant, but less so for loudness.

Table 4A shows that exponents for RE(BP) were remarkably consistent across conditions. For the RE(C) model, however, directional trends in exponents were mostly consistent with prior research: for both loudness and heaviness, the exponent for the middle standard was higher; for loudness, multiple judgments resulted in the higher exponent. These trends were observed for most of the subjects. However, for heaviness, the RE(C) exponent was higher for fractional judgments.

Results on the stability of exponents are presented in Table 4C. In all cases the correlations are elevated by use of the RE(BP) model, more so in the case of loudness. The heaviness correlations for the RE(C) model are surprisingly high, reflecting the weaker bias effect for heaviness compared to loudness.⁶ This result is consistent with the weaker effect sizes for heaviness reported in Table 4B. Nevertheless, the RE(BP) model does about as much better as can be expected considering the base-rate phenomenon, with all correlations above .90.

Table 4B also shows that effect size for the RE(C) model was substantial for loudness, with partial ES_c reaching 40% for response type, and 62% for standards. Effect size was smaller for heaviness, but still large (partial ES_c of 41% for response type and 20% for standards). On the other hand, for the RE(BP) model, even the partial ES_c values were zero. It is clear that the bias model (Equation 5) has eliminated the effect of standards and response type observed when bias is not taken into account.

With respect to the discussion in the previous section on differences in familiarity with the stimulus distribution, the very poor results for the RE(C) model suggest very strongly that it is the introduction of bias parameters into the RE power function that accounts for the more favorable results for RE (compared to ME), not greater familiarity with the stimulus distribution, since the results for the RE(C) model—not taking account of bias—were much worse than for ME in goodness-of-fit and effect size.

DISCUSSION

To put our results in perspective, we have extended the analyses of previous studies on the effect of standards by estimating the effect size and interpreting the bias effect in light of subject effects and the stability of exponents between levels of the independent variable. By so doing, we are able to conclude from Tables 2 and 3 that the RE power function (Equation 7), derived from the ratio of ratios condition (Equation 6), provides estimates of exponents that are relatively uninfluenced by the effects of standards: effect size is very low in all cases, and the correlations of exponents between conditions is relatively high. The effect of subjects for RE is powerful, however, accounting for about two thirds of the variability in exponents. The fact that for RE the coefficient of identity is mostly only slightly below Pearson r signifies that the

high correlations are not hiding some linear change in context, and that the e' values measure the degree to which the exponents are *identical*, relative to changes in context.

The results in Tables 2 and 3 also suggest that the effects of equal-range segments on estimation of exponents resulting from ME are much more serious than the effects of response type and standards are for RE. Although the ES values for ME are also low, bias condition is significant in two of the four cases for ME, and in none of the cases for RE (Table 2). Furthermore, the correlations between exponents for ME are quite low relative to those for RE, except for heaviness under Conditions C1 and D1 (Table 3). On the other hand, the subject effect is much smaller for ME.

The comparisons of the RE(C) and RE(BP) models (Table 4) show that when bias is ignored, estimates of exponents based on RE show strong directional trends consistent with previous research, but the introduction of bias parameters virtually eliminates systematic influences of standards and response type on the exponents.

Two studies on the stability of exponents are of particular interest in relation to our findings, those of Logue (1976) and M. Teghtsoonian and R. Teghtsoonian (1983). Logue reported that for ME of loudness, the correlation between subjects' exponents after an 11-week delay was .59 and statistically significant. Teghtsoonian and Teghtsoonian also found that ME of loudness yielded significant correlations, in this case after a delay of 1 week.

Whereas these two studies reported reliable correlations for ME of loudness after long delays, we see from Table 3 that the loudness correlations between equal-range segments for ME were mostly unreliable. Still, it is not safe to conclude that exponents for ME of loudness are more unstable between equal-range segments than between sessions with fairly long temporal separation (but within context), since the number of stimuli on which the exponent estimates are based are much smaller for the segment conditions.

M. Teghtsoonian and R. Teghtsoonian (1983) also reported that the correlation between individual exponents for loudness obtained in successive judgment sessions was not reliable when a standard was used and the modulus was changed. This result appears to be inconsistent with the positive results obtained for RE in the present study (Tables 3C and 4C). The critical difference between the two cases is that in the present study, we used power functions (Equations 5 and 7) that were constructed to "remove" the biasing effects of standards from the estimation of the exponents. To the extent that the model is a good approximation, the ES values for bias condition should be low, which they are (Tables 2A and 4B), and the correlations of exponents for bias condition should be, and are, relatively high (Tables 3C and 4C). The Teghtsoonian and Teghtsoonian procedure for estimating exponents, however, allowed (intentionally) the biasing effect of moduli to affect the magnitude of the exponents. Hence the results are not inconsistent, since different power functions were used in the two cases. Rather, since

Teghtsoonian and Teghtsoonian found that the bias induced by a change of modulus within the same session reduced the correlation to a nonsignificant level, we may take the two results conjointly as indicating support of the effectiveness of Equations 5 and 7 in providing bias-free estimates of the exponents. Whether exponents estimated from Equations 5 and 7 are stable over time remains to be seen.

The stability of exponents has also been studied between modalities. Positive correlations have been found among individual exponents from different continua (Ekman, Hosman, Lindman, Ljungberg, & Åkesson, 1968; Jones & Marcus, 1961; Rule, 1966), and it has been suggested that individual differences in the use of numbers could be responsible for these correlations (see Baird & Noma, 1978, p. 102; Rule & Markley, 1971). For the present data, the Pearson r was computed between loudness and heaviness for RE and ME, but neither of the coefficients was significantly different from zero (Table 3B, $p > .05$). However, the size of the correlations are comparable to those found in earlier studies, and given the small N , we do not view this result as a failure to replicate the finding on the salience of response factors.

Although we interpret our data as generally supportive of the bias form of the RE power function, applied to ratio judgments, the generality of the power function is seriously in question. Numerous studies point to the wide subject variability in exponents for a single modality, and the exponents vary widely under a variety of experimental conditions (Poulton, 1968). Birnbaum and Elmasian (1977) concluded that, even under the assumption of a ratio model, their loudness data were incompatible with a power function, a conclusion reached also by Schneider, Parker, Farrell, and Kanow (1976). Whether the bias model given by Equation 5 could account for their data is an open question.

An alternative approach to complicating the simple "quantitative" power function through, for example, the introduction of bias parameters, is to focus on "qualitative" psychophysical laws. This approach is consistent with that of Birnbaum and associates (see Birnbaum, 1982) and of Schneider and associates (Schneider et al., 1976; Schneider, Parker, & Stein, 1974) who reached strong conclusions based on ordinal data.

More recently, Adams and Fagot (1987) have formulated and tested qualitative laws such as monotonicity and convexity (that psychophysical functions are upwardly or downwardly convex). Adams and Fagot also propose a study of "psychophysical invariants": given the lack of strict invariance of the exponent of the power function, one should search for those properties of a psychophysical function that are invariant. For example, it appears that in some cases, the direction of convexity is invariant. Thus Marks (1974a) reported exponents for loudness based on 40 studies, using 13 different scaling methods, and in each of the studies the direction of convexity was downward, although the exponents varied widely from

0.13 to 0.85. The direction of convexity also appears to be invariant downward across subjects for brightness (Marks & J. C. Stevens, 1966); and invariant upwards for handgrip (J. C. Stevens & Mack, 1959), muscular effort (Bernyer, 1962), and saltiness and sweetness (Ekman & Åkesson, 1965). Presumably simpler sensory-neural models would be sufficient to explain such weaker invariances. Such qualitative approaches should be intended to complement, not replace, quantitative models.

REFERENCES

- ADAMS, E. W., & FAGOT, R. F. (1987). Monotonicity, convexity, and other qualitative psychophysical laws. *Journal of Mathematical Psychology*, *31*, 113-134.
- BAIRD, J. C., & NOMA, E. (1978). *Fundamentals of scaling and psychophysics*. New York: Wiley.
- BECK, J., & SHAW, W. A. (1961). The scaling of pitch by the method of magnitude estimation. *American Journal of Psychology*, *74*, 242-251.
- BECK, J., & SHAW, W. A. (1965). Magnitude of the standard, numerical value of the standard, and stimulus spacing in the estimation of loudness. *Perceptual & Motor Skills*, *21*, 151-156.
- BERNYER, G. (1962). Étude sur la Validité d'une Échelle de Sensation d'effort Musculaire. *Année Psychologique*, *62*, 1-15.
- BIRNBAUM, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401-485). Hillsdale, NJ: Erlbaum.
- BIRNBAUM, M. H., & ELMASIAN, R. (1977). Loudness "ratios" and "differences" involve the same psychophysical operation. *Perception & Psychophysics*, *22*, 383-391.
- COHEN, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational & Psychological Measurements*, *33*, 107-112.
- CROSS, D. V. (1973). Sequential dependencies and regression in psychophysical judgments. *Perception & Psychophysics*, *14*, 547-552.
- CROSS, D. V. (1982). On judgments of magnitude. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 73-88). Hillsdale, NJ: Erlbaum.
- DODD, D. H., & SCHULTZ, R. F., JR. (1973). Computational procedures for estimating magnitude of effect for some analysis of variance designs. *Psychological Bulletin*, *79*, 391-395.
- EKMAN, G., & ÅKESSON, C. (1965). Saltiness, sweetness, and preference: A study of qualitative relations in individual subjects. *Scandinavian Journal of Psychology*, *6*, 241-253.
- EKMAN, G., HOSMAN, B., LINDMAN, R., LJUNGBERG, L., & ÅKESSON, C. (1968). Interindividual differences in scaling performance. *Perceptual & Motor Skills*, *26*, 815-823.
- FAGOT, R. F. (1978). A theory of relative judgment. *Perception & Psychophysics*, *24*, 243-252.
- FAGOT, R. F. (1981). A theory of bidirectional judgments. *Perception & Psychophysics*, *30*, 181-193.
- FAGOT, R. F. (1982). Constructability and the power function for line judgments. *Perception & Psychophysics*, *31*, 392-394.
- HELLMAN, R. P., & ZWISLOCKI, J. J. (1961). Some factors affecting the estimation of loudness. *Journal of the Acoustical Society of America*, *33*, 687-694.
- JONES, F. N., & MARCUS, M. J. (1961). The subject effect in judgments of subjective magnitude. *Journal of Experimental Psychology*, *61*, 40-44.
- KEPPEL, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- KRANTZ, D. H. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, *9*, 168-199.
- LOGUE, A. W. (1976). Individual differences in magnitude estimation of loudness. *Perception & Psychophysics*, *19*, 279-280.

- MARKS, L. E. (1974a). On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as science. *Perception & Psychophysics*, **16**, 358-376.
- MARKS, L. E. (1974b). *Sensory processes: The new psychophysics*. New York: Academic Press.
- MARKS, L. E., & CAIN, W. S. (1972). Perception of intervals and magnitudes for three prothetic continua. *Journal of Experimental Psychology*, **94**, 6-17.
- MARKS, L. E., & STEVENS, J. C. (1966). Individual brightness functions. *Perception & Psychophysics*, **1**, 17-24.
- MYERS, J. L. (1979). *Fundamentals of experimental design*. Boston: Allyn & Bacon.
- PARDUCCI, A. (1982). Category ratings: Still more contextual effects! In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 89-105). Hillsdale, NJ: Erlbaum.
- POULTON, E. C. (1968). The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, **69**, 1-19.
- POULTON, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, **86**, 777-803.
- RULE, S. J. (1966). Subject differences in exponents of psychophysical power functions. *Perceptual & Motor Skills*, **23**, 1125-1126.
- RULE, S. J., & MARKLEY, R. P. (1971). Subject differences in cross modality matching. *Perception & Psychophysics*, **9**, 115-117.
- SCHNEIDER, B., PARKER, S., FARRELL, G., & KANOW, G. (1976). The perceptual basis of loudness ratio judgments. *Perception & Psychophysics*, **19**, 309-320.
- SCHNEIDER, B., PARKER, S., & STEIN, D. (1974). The measurement of loudness using direct comparisons of sensory intervals. *Journal of Mathematical Psychology*, **11**, 259-273.
- SHEPARD, R. N. (1981). Psychological relations and psychological scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology*, **24**, 21-57.
- STEVENS, J. C., & GUIRAO, M. (1964). Individual loudness functions. *Journal of the Acoustical Society of America*, **36**, 210-213.
- STEVENS, J. C., & MACK, J. D. (1959). Scales of apparent force. *Journal of Experimental Psychology*, **58**, 405-413.
- STEVENS, J. C., & TULVING, E. (1957). Estimation of loudness by a group of untrained observers. *American Journal of Psychology*, **70**, 600-605.
- STEVENS, S. S. (1956). The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, **69**, 1-25.
- STEVENS, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153-181.
- STEVENS, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.
- SUSSKIND, E. C., & HOWLAND, E. W. (1980). Measuring effect magnitude in repeated measures ANOVA designs: Implications for gerontological research. *Journal of Gerontology*, **35**, 867-876.
- TEGHTSOONIAN, M., & TEGHTSOONIAN, R. (1983). Consistency of individual exponents in cross-modal matching. *Perception & Psychophysics*, **33**, 203-214.
- 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.
- WARD, L. M. (1986). Mixed-modality psychophysical scaling: Double cross-modality matching for "difficult" continua. *Perception & Psychophysics*, **39**, 407-417.
- ZEGERS, F. E. (1986). A family of chance-corrected association coefficients for metric scales. *Psychometrika*, **51**, 559-562.
- ZEGERS, F. E., & TEN BERGE, J. M. F. (1985). A family of association coefficients for metric scales. *Psychometrika*, **50**, 17-24.

NOTES

1. The derivation of scales in which parameters cancel may also be accomplished by "mixed-modality scaling" (see Ward, 1986).

2. As described under Method, the subjects did not complete blocks of ME trials for successive equal-range segments, but were presented with the entire set of nine stimuli in each block of trials. Hence equal-range segment does not constitute a "context" in the sense that standards do for RE. Nevertheless, if the ME power function hypothesis is correct, then the exponents should be invariant relative to the location of the equal-range segment (e.g., low vs. high subsets of stimuli), for either method of presentation. But because of this difference, we shall use the more neutral term *condition* instead of "context," and *bias effects* instead of "context effects," for both RE and ME.

3. If additivity is not satisfied, the estimates of ES in Table 2 will be overestimates. Applying the Tukey single df test for additivity, nonadditivity was significant only in the case of ME/loud (B2) ($p < .025$). Assuming nonadditivity, ES_c for ME/loud (B2) was reduced from .28 to .24, the maximum reduction for the balance of the cases was .02, and the mean reduction for all eight cases was .01.

4. One argument made against estimating ES_c for the fixed effect on the basis of total variability in a repeated measures design is that the large subject variability overwhelms the fixed effect variability. This argument has been extended to other multifactor designs as well, for which reason a number of authors advocate "partial" ES measures (Cohen, 1973; Keppel, 1982), defined as the variability due to an effect A as a proportion of the variability due to A plus error variability (i.e., "partialling out" all other nonerror sources of variability). This means that the partial ES_c estimates the proportion of within-subject variability accounted for by bias condition, and the partial ES_s estimates the proportion of between-subject variability accounted for by subjects.

Table 2 shows the partial ES values in parentheses following the ES values based on total variability. For example, A2 indicates that for REH, 2% of the total variability, but 10% of the within-subject variability, is accounted for by standards. The differences are not great: for RE, ES_c increases from a mean of .02 based on total variability to .07 for within-subject variability; and for ME, from .11 to .15.

5. The Zegers (1986) coefficient of identity (corrected for chance) is

$$e' = \frac{2(\Sigma XY - (1/N)\Sigma X \Sigma Y)}{\Sigma X^2 + \Sigma Y^2 - 2(1/N)\Sigma X \Sigma Y}$$

6. We note that the heaviness correlations are actually higher than those reported for RE in Table 3. Our only explanation for this lies in reliability differences: the dependent variable for the RE(C) regression is $Y = \ln R_{ax}$ and for RE (Equation 7) it is $Y = \ln R_{ax} - \ln R_{bx}$. So presumably, the difference is due to the greater unreliability of difference scores.

(Manuscript received August 17, 1987;
revision accepted for publication September 7, 1988.)