Journal of Experimental Psychology: General
1990, Vol. 119, No. 4, 375–396

0096-3445/90/$00.75

Sequential Effects in Magnitude Scaling: Models and Theory

Lawrence T. DeCarlo and David V. Cross
State University of New York at Stony Brook

Research on sequential effects in magnitude scaling is reviewed, and its implications about the adequacy of current time series regression models is discussed. A regression model that unifies what at first appear to be contradictory results is proposed. Theoretical models of judgment and perception are introduced, and their relation to alternative regression models is clarified. A theoretical model of relative judgment that clarifies the role of judgmental error and frames of reference in magnitude scaling is examined in detail. Four experiments that test the model are presented. The results, along with recent results presented by Ward (1987), provide support for the model. The importance of being explicit about the relation of theoretical models to regression models and about the role of error in these models is discussed.

In the 1950s, Stevens (e.g., 1956, 1957) popularized a new class of procedures where subjects "directly" indicated sensation magnitude by responding to presented magnitudes of physical stimuli with numbers (or stimuli). Stevens observed that, when plotted on log-log coordinates, the (geometric) means of responses given to each stimulus intensity showed a linear increase. This led him to conclude that response magnitude, R, is a power function of stimulus magnitude, S:

$$R = \alpha S^{\beta} \tag{1}$$

where α and β are parameters. Stevens also assumed that subjects' responses in magnitude scaling experiments are proportional to sensation magnitude, Ψ:

$$R = \alpha \Psi \tag{2}$$

Stevens used the above assumption, along with the observed relation between responses and stimulus intensity (Equation 1), to argue that the *psychophysical function*, which relates unidimensional sensation magnitude to stimulus intensity, is a power function:

$$\Psi = S^{\beta} \tag{3}$$

where β, in Stevens's view (e.g., 1975), is a parameter that characterizes the particular sensory continuum under investigation. Although the empirical relation of Equation 1 is of interest in and of itself, it has received considerable attention in psychophysics largely because of its assumed relation to Equation 3. In particular, if Equation 2 is correct, then Equation 1 provides information about the form of the psychophysical function and an estimate of its exponent.

Research has shown, however, that Equation 1 is an incomplete empirical model, because there are systematic variations in the data that it fails to account for. This is particularly evident in research on sequential effects in magnitude scaling (e.g., Cross, 1973; Luce & Green, 1974a; Ward, 1973, 1979), which has shown that subjects' responses show dependencies over time that are not accounted for by Equation 1. In response, several researchers have proposed regression models that are generalizations of Equation 1. The present article focuses on these models and on the theories underlying them. The first section of the article presents a selective review of empirical findings that have direct implications about the adequacy of alternative regression models. Current regression models are reviewed, and difficulties associated with the interpretation of their parameters are discussed. A regression model that provides a unified account of what at first appear to be contradictory results is introduced. Next, theoretical models of judgment and perception are introduced, and their relation to alternative regression models is clarified. Lastly, experiments suggested by one of the judgmental theories are presented.

Sequential Effects

One of the reasons it is difficult to arrive at general conclusions about sequential effects is because, as Staddon, King, and Lockhead (1980) have noted, there is no agreed-upon method of analysis. Nevertheless, this section shows that research in the area has established several basic results that have direct implications about the adequacy of the regression models discussed below.

The standard approach for fitting Equation 1 is to first linearize it using logarithms and to then perform least squares regression. The model fit to the data, therefore, is:

$$\log R_i = \beta_0 + \beta_1 \log S_i + e_i \tag{4}$$

where $\beta_0 = \log \alpha$. The error term, e_i, is a random variable that represents the net effect of all variables not included in the model, which nevertheless influence responses. This omission occurs because, for example, it may not be known what all the relevant variables are or because relevant variables are not measurable.

Equation 4 is a *static* model, in that the response on each trial depends solely on the stimulus intensity and "noise" of that trial. Research on magnitude scaling, however, has shown

This article is based on a doctoral dissertation by Lawrence T. DeCarlo, submitted to the State University of New York at Stony Brook.

Correspondence concerning this article should be addressed to Lawrence T. DeCarlo, who is now at the Division of Biostatistics, School of Public Health, Columbia University, 600 West 168th Street, New York, New York 10032.

that the model should include a *dynamic* element, because there appear to be effects that extend over trials. Interest in these sequential effects follows from earlier demonstrations of their presence in identification and categorization experiments (e.g., Garner, 1953; Holland & Lockhead, 1968; Ward, 1972; Ward & Lockhead, 1970, 1971). For example, Garner (1953) showed, in an identification experiment, that the response, R_i, to stimulus intensity, S_i, tended to be larger when the immediately preceding stimulus intensity was greater than S_i and smaller when the preceding stimulus intensity was less than S_i. This was demonstrated by plotting, separately for each stimulus intensity, the mean response to S_i given the stimulus intensity presented on the immediately preceding trial, $S_{j,t-1}$. These plots showed a positive trend. The result is usually referred to as an "assimilation" of the current response towards the preceding stimulus intensity. Garner noted that it is not possible to determine whether the effect is due to the previous stimulus or to the previous response, because of the high correlation between these variables.

A similar result has been found in magnitude estimation and cross-modality matching experiments (e.g., Cross, 1973; Luce & Green, 1974a, 1974b; Ward, 1973, 1975). These studies have shown, using the plots discussed above, that the conditional expectation of log responses, given the current and previous log stimulus intensities, increases with the log stimulus intensity of the previous trial. That is, $E(\log R_i | \log S_i, \log S_{j,t-1})$ increases with $\log S_{j,t-1}$.[1]

Sequential effects have also been studied using correlations. For example, several studies (e.g., Baird, Green, & Luce, 1980; Green, Luce, & Duncan, 1977; Green, Luce, & Smith, 1980; Ward, 1975) have shown that successive log responses are positively correlated. The magnitude of the correlation also appears to depend on the similarity of successive stimuli. In particular, the correlations tend to be large when the difference between successive log stimuli is small and close to zero when the difference is large (see Jesteadt, Luce, & Green, 1977). This leads to an "inverted-V" pattern of correlations that is discussed below.

A result related but not identical to the finding of correlated log responses has been presented by Jesteadt et al., (1977), and by Ward (1979). These authors examined the coefficient β obtained for the following regression:

$$\log(R_t/\kappa S_t^{\gamma}) = \beta \log(R_{t-1}/\kappa S_{t-1}^{\gamma}) + \delta + \epsilon_t$$

where κ and γ were first estimated by fitting Equation 4 to each subject's data. It follows (with $\beta_0 = \log \kappa$ and $\beta_1 = \gamma$) that the terms in parentheses are the *residuals*, $\hat{e}_t$, for Equation 4. The above equation, therefore, represents a regression of $\hat{e}_t$ on $\hat{e}_{t-1}$, and the coefficient β measures the residual *autocorrelation*. The finding of autocorrelated residuals has different implications than the finding of autocorrelated log responses, as shown in the next section.

In summary, three results found in research on sequential effects are: (a) the mean response to log stimulus intensity, i, given the log stimulus intensity presented on the previous trial, $\log S_{j,t-1}$, increases with $\log S_{j,t-1}$; (b) successive log responses are positively correlated; and (c) successive residuals from Equation 4 are positively correlated. In the next section, we review regression models of sequential effects and consider the implications of the above three results for these models.

Time Series Regression Models

Although it is apparent from the above that the method of analysis has varied, most researchers have considered sequential effects as arising because of the influence of previous stimuli and/or responses. As a result, Equation 4 has been generalized by including previous stimuli and/or responses as regressors. For example, Cross (1973) observed assimilation to the previous stimulus (using the plots discussed above), and introduced the following regression model to account for this effect:

$$R_t = \alpha S_t^{\beta_1} (S_{t-1}/S_t)^{\beta_2} \epsilon_t$$

The subscript i of Equation 4 has been replaced by t because it is important to recognize the temporal order of responses to obtain information about underlying processes. Thus, all the models considered in this article are *time series* regression models.

Rearranging terms and taking logarithms gives:

$$\log R_t = \beta_0 + (\beta_1 - \beta_2) \log S_t + \beta_2 \log S_{t-1} + e_t \quad (5)$$

where $\beta_0 = \log \alpha$ and $e_t = \log \epsilon_t$. Equation 5 shows that, if β_2 is positive, then the exponent of Stevens' power law underestimates the "true" exponent, β_1, by β_2. That is, a least squares fit of Equation 4, which omits $\log S_{t-1}$, provides an unbiased estimate of $\beta_1 - \beta_2$, and not simply of β_1.[2] Cross suggested that this underestimation of the exponent may account for the "regression effect" frequently found in psychophysical research (see Cross, 1973; Stevens & Greenbaum, 1966). Equation 5 is also related to an explanation of sequential effects (for identification experiments) offered by Garner (1953): "the response to a stimulus is actually a response to the weighted mean of the present stimulus and that heard previously" (p. 379). That is:

$$\log R_t = \beta_0 + (1 - \lambda)\beta_1 \log S_t + \lambda\beta_1 \log S_{t-1} + e_t$$

where λ represents the weighting factor. The above is equivalent to Equation 5, with $\lambda = \beta_2/\beta_1$.

Cross (1973) fit Equation 5 to data from a magnitude estimation experiment where the loudness of noise was judged; he found assimilation to the previous log stimulus (i.e., the estimate of β_2, 0.055) was positive. Cross also obtained a positive $\hat{\beta}_2$ (0.177) for a fit of Ward's 1973 data. Jesteadt et al. (1977) fit Equation 5 to their data and obtained a positive $\hat{\beta}_2$ (0.100 for sound pressure; the reported value of 0.050 was for sound power).

Taking conditional expectations of Equation 5 gives:

$$E(\log R_i | \log S_i, \log S_{j,t-1}) = \beta_0 + \beta_1 \log S_i + \beta_2 \log S_{j,t-1}$$

[it follows from the assumptions about the error term that $E(e_i | \log S_i, \log S_{j,t-1}) = 0$]. The above equation shows that Equation 5, with positive β_2, is consistent with the assimilation

[1] Luce and Green (1974a) examined plots of the above conditional expectation divided by $E(\log R_i | \log S_i)$ (see their Equation 9).

[2] The least squares estimate of $(\beta_1 - \beta_2)$ is unbiased because the omitted variable, $\log S_{t-1}$ is, by design, orthogonal to S_t (see Rao, 1971).

observed in the plots discussed above (Result 1); the model was in fact proposed to account for this result. Equation 5 is also consistent with the finding of autocorrelated log responses (Result 2); writing out the equation for log R_t and log R_{t-1} shows that the current and previous log response are both determined in part by log S_{t-1}. However, Equation 5 is not consistent with the finding that Equation 4's residuals are autocorrelated (Result 3). To see why, suppose for the moment that Equation 5 is correct. Then, if Equation 4 is fit instead, $e_t = \beta_2 \log S_{t-1} + u_t$, where u_t represents random error. It follows that e_t and e_{t-1} are not correlated, because log S_{t-1} and log S_{t-2} are, by design, uncorrelated. Equation 5, therefore, is not complete; although it accounts for Results 1 and 2, it is not consistent with Result 3. This is one situation where the autocorrelation of residuals provides different information than the autocorrelation of log responses.

Jesteadt et al. (1977) further generalized Equation 5 by including log R_{t-1} as a regressor:

$$\log R_t = \beta_0 + \beta_1 \log S_t + \beta_2 \log S_{t-1} + \beta_3 \log R_{t-1} + e_t \quad (6)$$

(where $\beta_1 - \beta_2$ has been replaced by β_1 to simplify the notation). Empirical evidence for this generalization comes from research showing slight increases in R^2 when log R_{t-1} is included in the model along with log S_{t-1} (Green, Luce, & Duncan, 1977; Ward, 1979, 1987). Jesteadt et al. (1977) examined increments in R^2 obtained when either previous stimuli or previous responses were included in the model, but not both.

Equation 6 was also motivated by the *response ratio hypothesis* of Luce and Green (1974a, 1974b), which consists of two ideas. The first is that response ratios are proportional to ratios of internal representations. Luce and Green (1974a) noted that this idea is consistent with the instructions usually given in magnitude estimation experiments and with evidence provided by plots of response ratios against stimulus ratios. The second idea is that each stimulus presentation gives rise to (at least) two independent representations, Ψ_t and Ψ_t^*, where Ψ_t^* is the representation used for comparative judgment on trial t + 1, and $\Psi_t^* \neq \Psi_t$. Combining these two ideas gives:

$$\frac{R_t}{R_{t-1}} = \alpha \frac{\Psi_t}{\Psi_{t-1}^*} \quad (7)$$

When $\alpha \neq 1$, Equation 7 predicts that log responses will show a linear trend over trials (as shown in the next paragraph).

Luce and Green (1974a, 1974b) proposed the response ratio hypothesis to account for Result 1, and for a possible (linear) drift of responses over time. The assumption of two independent representations was considered necessary because

$$\frac{R_t}{R_{t-1}} = \alpha \frac{\Psi_t}{\Psi_{t-1}}$$

reduces to:

$$R_t = \alpha^{t-1}\Psi_t(R_1/\Psi_1)$$

and the above equation does not account for sequential effects (the term α^{t-1} shows why the above, in log form, predicts that log responses will show a linear trend over trials). It should be noted, however, that if an error term is included in the model, then

$$\frac{R_t}{R_{t-1}} = \alpha \frac{\Psi_t}{\Psi_{t-1}}\epsilon_t$$

where ϵ_t represents judgmental error, reduces to

$$R_t = \alpha^{t-1}{}_{\Psi t}(R_1/\Psi_1)\Pi_{i=2}^{t}\epsilon_i$$

and the above equation predicts a correlation between successive responses, because R_t and R_{t-1} are both determined in part by ϵ_{t-1}, ϵ_{t-2}, and so on (this can be seen by writing out the equation for R_t and R_{t-1}). Thus, the hypothesis predicts sequential effects (the autocorrelation of responses and residuals, Results 2 and 3) when an error term is introduced, as was done by Marley (1976) in his revision of the response ratio hypothesis (see his Equation 2). Luce and Green (1974a) considered another approach—the assumption of independent representations—and showed that it accounts for Result 1 (see their Equation 9 and Figure 1).

The response ratio hypothesis is a theoretical model that involves unobservable sensation magnitudes, Ψ. To determine its relation to the empirical model of Equation 6, assumptions about the relation between sensation magnitude and stimulus magnitude must be made. For example, if it is assumed that $\Psi_t = S_t^{\beta_1}$ and $\Psi_{t-1}^* = S_{t-1}^{\beta_2}$ (in Cross's 1973 approach, β_1 is replaced by $\beta_1 - \beta_2$), then substituting in Equation 7, taking logarithms, and rearranging gives:

$$\log R_t = \beta_0 + \beta_1 \log S_t - \beta_2 \log S_{t-1} + \log R_{t-1}$$

where $\beta_0 = \log \alpha$. The above model can be fit using least squares by regressing log response ratios, log (R_t/R_{t-1}), on log S_t and log S_{t-1} (i.e., log R_{t-1} can be brought to the left side of the equation). Jesteadt et al.'s (1977) approach, however, was to fit the more general model of Equation 6 (it has one additional parameter, β_3). In this case, the estimates of the coefficient of log R_{t-1} (β_3) were considerably less than unity (0.50 or less). Because the response ratio hypothesis does not predict this result, Jesteadt et al. (1977) concluded that the hypothesis, as it stands, is not correct. Luce, Baird, Green, and Smith (1980) have discussed alternatives.

Although Jesteadt et al. (1977) rejected the (unmodified) response ratio hypothesis, researchers have continued to use Equation 6 to study sequential effects (e.g., Jesteadt et al., 1977; Ward, 1979, 1987). The model is clearly consistent with Results 2 and 3 discussed above—the inclusion of log R_{t-1} in the model accounts for the autocorrelation. The usual interpretation of Result 1, however, poses something of a problem. Result 1 has often been interpreted as showing that the previous stimulus exerts an assimilative influence on the current response. This interpretation is also consistent with the positive estimates of the coefficient of log S_{t-1} obtained for Equation 5. On the other hand, the results for Equation 6 are the opposite—the estimates of the coefficient of log S_{t-1} are consistently negative, which suggests contrast, not assimilation, to the previous stimulus, For example, previous research has shown that fits of Equation 5 to data from Cross (1973), Jesteadt et al. (1977), and Ward (1973) gave estimates of the coefficient of log S_{t-1} equal to 0.055, 0.177, and 0.100

(for sound pressure), respectively, whereas fits of Equation 6 to the same data gave estimates of −0.026, −0.111, and −0.104. Although these results appear to be contradictory, they have received little attention.

Fortunately, a different view of Equation 6 arises when its relation to another empirical model is considered. In particular, another way of generalizing Equation 4 is by including a *first-order autoregressive* [AR(1)] error process:

$$e_t = \rho e_{t-1} + u_t \quad (8)$$

where ρ is referred to as the *autocorrelation parameter*, and u_t is a random variable with zero mean, constant variance, no correlation with its previous values, and no correlation with e_{t-1} (or log S_t in Equation 9 below). The restriction $|\rho| < 1$ ensures that the error process is *stationary*; that is, it has a finite mean and variance, and a covariance structure that does not depend on time (e.g., see Johnston, 1984).

Combining Equation 8 with Equation 4 gives:

$$\log R_t = \beta_0 + \beta_1 \log S_t + \rho e_{t-1} + u_t, \quad (9)$$

which is simply Stevens's power law, in log-linear form, with an AR(1) error process. The usual interpretation of autocorrelated errors is that a variable (or variables) that has effects extending over more than one time period has been omitted from the model (e.g., see Johnston, 1984; Kmenta, 1986). It is easy to think of a number of omitted factors that might influence subjects' responses in magnitude scaling experiments: the subject's level of attention or motivation, memory, the strategy or strategies the subject uses to arrive at a response, and so on. Changes in these factors over the course of the experiment can lead to autocorrelated errors. Of course, if the relevant omitted variable (or variables) is known and is also measurable, it should be included in the model. However, in situations where the omitted variable(s) cannot be measured and/or experimentally controlled, Equation 9 may be appropriate.

The relation of Equation 9 to Equation 6 can be shown as follows. According to Equations 8 and 9,

$$e_t = \rho e_{t-1} + u_t = \log R_t - \beta_0 - \beta_1 \log S_t.$$

Because this relation also holds on trial t − 1, it follows that $e_{t-1} = \log R_{t-1} - \beta_0 - \beta_1 \log S_{t-1}$. Substituting this expression for e_{t-1} in Equation 8 gives:

$$\rho e_{t-1} + u_t = \rho \log R_{t-1} - \rho\beta_0 - \rho\beta_1 \log S_{t-1} + u_t.$$

Substituting this expression for $\rho e_{t-1} + u_t$ in Equation 9 gives

$$\log R_t = (1 - \rho)\beta_0 + \beta_1 \log S_t - \rho\beta_1 \log S_{t-1} + \rho \log R_{t-1} + u_t. \quad (10)$$

The above shows that Equation 9 is equivalent to Equation 6, with $\rho = \beta_3$ and $\beta_2 = -\rho\beta_1$. Thus, Equation 10 shows that, if the error process is stationary (i.e., $|\rho| < 1$), then the estimate of β_3 obtained for Equation 6 will be less than unity, which is exactly what Jesteadt et al. (1977) found. In addition, the coefficients of log R_{t-1} ($\beta_3 = \rho$) and log S_{t-1} ($-\rho\beta_1$) will have opposite signs, provided that β_1 is positive, which is usually the case. This means that the "contrast" found for Equation 6 might appear simply because the errors of Equation 4 are autocorrelated. Clearly, in the absence of an explicit theory, Equation 6's coefficients must be interpreted with caution (see the Theory section below).

Equations 9 and 10 are also relevant to Ward's (1979) claim that the findings $0 < \beta_3 < 1$ and $|\beta_2| < \beta_1$ for Equation 6 support his fuzzy judgment model over the response ratio hypothesis. In light of Equation 10, $|\beta_3| < 1$ indicates that the error process is stationary (Equation 10 shows that $\beta_3 = \rho$). In addition, $|\beta_3| < 1$ together with the parameter restriction implied by Equation 9, $\beta_2 = -\beta_1\beta_3$, yield Ward's second result, $|\beta_2| = |\beta_1\beta_3| < \beta_1$. The two results noted by Ward (1979), therefore, will be found if the errors of Equation 4 are autocorrelated (autocorrelation is more the rule than exception for time series data). For this reason, they do not provide particularly compelling support for the fuzzy judgment model.

In sum, Equations 9 and 10 show that (a) the finding that $\beta_3 < 1$ for Equation 6 is consistent with a stationary error process in Equation 9, (b) the finding of a negative β_2 for Equation 6 is consistent with positive autocorrelation in Equation 9, and (c) the finding that $|\beta_2| < \beta_1$ for Equation 6 is consistent with the parameter restriction implied by Equation 9.

As was noted above, the results obtained for Equation 6 are contradictory to those obtained for Equation 5. The next section shows that Equation 5, Equation 6, and Result 1 are unified by an alternative empirical model.

An Alternative Model

In this section, we consider the possibility that sequential effects arise not because of the influence of the previous stimulus and previous response, as suggested by Equation 6, but because of the influence of the previous stimulus and an autocorrelated omitted variable(s) (attention, memory, strategy, motivation, etc.; a specific interpretation of autocorrelation is presented in the Theory section below). That is, log R_{t-1} in Equation 6 may be serving as a proxy for some other autocorrelated omitted variable(s). An alternative to Equation 6, therefore, is:

$$\log R_t = \beta_0 + \beta_1 \log S_t + \beta_2 \log S_{t-1} + \rho e_{t-1} + u_t \quad (11)$$

(DeCarlo, 1989, 1989/1990). The above equation can be rewritten following the steps used to derive Equation 10 from Equation 9. This gives:

$$\log R_t = \beta_1 \log S_t + (\beta_2 - \rho\beta_1) \log S_{t-1} + \rho \log R_{t-1} - \rho\beta_2 \log S_{t-2} + u_t \quad (12)$$

(the intercept has been dropped). It is important to recognize that, when log S_{t-2} is omitted, the above equation is equivalent to Equation 6 (with different parameters). Equation 12 shows, therefore, that a least squares fit of Equation 6 will yield unbiased estimates of β_1 and $\beta_2 - \rho\beta_1$; the estimate of ρ will be biased because log R_{t-1} is correlated with an omitted variable, namely, log S_{t-2}. However, it is shown below that the estimate of the coefficient of log R_{t-1} obtained for Equation 6 is close to the estimate of ρ obtained for Equation 11.

Equations 11 and 12 shed light on the seemingly contradictory results obtained for Equations 5 and 6, that is, the change in the sign of the coefficient of log S_{t-1} from positive to negative. Assume for the moment that Equation 11 is correct. In this case, fits of Equation 5 provide an unbiased estimate of β_2 in Equation 11 (neglecting the autoregressive error process and inappropriately using least squares will give unbiased estimates of β_1 and β_2, although they will no longer have minimum variance). Thus, the results obtained for Equation 5 suggest that β_2 in Equation 11 is small and positive. In addition, Equation 12 shows that the coefficient of log S_{t-1} for Equation 6 is equal to $\beta_2 - \rho\beta_1$. This means that if the product of (positive) ρ and β_1 is greater than β_2, then the coefficient of log S_{t-1} will be negative for Equation 6, even if β_2 in Equation 11 is positive. Equations 11 and 12 show, therefore, that a positive $\hat{\beta}_2$ for Equation 5 and a negative estimate for Equation 6 are not necessarily contradictory results.

Although Equation 11 has not been fit to published data, Equation 12 shows that an estimate of β_2 can be obtained by multiplying together $\hat{\beta}_1$ and $\hat{\beta}_3$ obtained for Equation 6 (they provide estimates of β_1 and ρ in Equation 11, respectively) and adding the result to the estimated coefficient of log S_{t-1}. The values obtained in this way for Cross's 1973 data (0.052), Ward's 1973 data (0.189), and Jesteadt et al.'s 1977 data (0.105) are close to those reported for fits of Equation 5 (0.055, 0.177, and 0.100, respectively), which suggests that the estimates are reasonably accurate.

Table 1 presents the average coefficients obtained for all (to our knowledge) published magnitude estimation and cross-modality matching experiments that have fit Equation 6. Table 2 presents the estimates of Equation 11's coefficients. The estimates of β_2 were obtained as described above; the estimates of β_1 and ρ (i.e., β_3) were taken from Table 1.

Table 1 shows that fits of Equation 6 give positive and negative estimates for the coefficients of log R_{t-1} and log S_{t-1}, respectively. This is the usual finding of "assimilation" to the previous response and "contrast" to the previous stimulus. Table 2, on the other hand, shows that the estimates of the coefficient of log S_{t-1} for Equation 11 are small but consistently positive. Table 2 shows, therefore, that the results obtained for Equation 6 are consistent with positive values of β_2 in Equation 11. This result is important because a positive β_2 in Equation 11 is consistent with the results found for Equation 5 and with the positive relation of Result 1. Equation 11, in other words, unifies Result 1, Cross's (1973) results, and Jesteadt et al.'s (1977) results.

Table 1
Coefficients for Equation 6 Obtained in Published Research

Task	Loudness	Study	S_t	S_{t-1}	R_{t-1}
ME		Cross (1973)[a]	.584	−.026	.133
ME	Ratio ME	Ward (1973)	.502	−.111	.598
ME	Ratio ME	JLG (1977)[b]	.546[d]	−.104[d]	.382
ME	Ratio ME	GLD (1977)[c]	.438[d]	−.174[d]	.531
ME	Absolute ME	Ward (1987)	.483	−.101	.391
ME	Ratio ME	Ward (1987)	.400	−.114	.544
CMM		Ward (1975)	.326	−.070	.269
CMM		Ward (1987)	.344	−.079	.332
	Distance				
ME	High Information	Ward (1979)	.931	−.109	.143
ME	Medium Information	Ward (1979)	.191	.011	.422
CMM	High Information	Ward (1979)	.539	−.126	.269
CMM	Medium Information	Ward (1979)	.381	−.078	.175

[a] Cross's 1973 results are from fits of the actual data. [b] Jesteadt, Luce, & Green, 1977. [c] Green, Luce, & Duncan, 1977. [d] Values of the coefficients for sound pressure (originally reported for sound power).

Table 2
Estimated Coefficients for Equation 11 for the Studies of Table 1

Task	Loudness	Study	S_t	S_{t-1}	e_{t-1}
ME		Cross (1973)	.584	.052	.133
ME	Ratio ME	Ward (1973)	.502	.189	.598
ME	Ratio ME	JLG (1977)	.546	.105	.382
ME	Ratio ME	GLD (1977)	.438	.059	.531
ME	Absolute ME	Ward (1987)	.483	.088	.391
ME	Ratio ME	Ward (1987)	.400	.104	.544
CMM		Ward (1975)	.326	.018	.269
CMM		Ward (1987)	.344	.035	.332
	Distance				
ME	High information	Ward (1979)	.931	.024	.143
ME	Medium information	Ward (1979)	.191	.092	.422
CMM	High information	Ward (1979)	.539	.019	.269
CMM	Medium information	Ward (1979)	.381	−.011	.175

In summary, although Equation 6 offers one approach to the study of sequential effects, Equation 11 presents an alternative empirical model that is worthy of examination. Equation 11 differs from Equation 6 in that it allows for the possibility that autocorrelation might arise because an unspecified variable has been omitted from the model. In addition, Equations 11 and 12, along with Tables 1 and 2, show that positive autocorrelation can lead to "contrast" in Equation 6 (i.e., a negative β_2) even if the previous stimulus actually exerts an "assimilative" influence (i.e., a positive β_2 in Equation 11).

Up to this point, we have focused on empirical models of sequential effects. Psychophysicists, however, as well as psychologists in general, are interested in the observable variables of regression models primarily because of the potential information they offer about unobservable psychological constructs. For example, as was pointed out in the introduction, Equation 1 has received attention in psychophysics largely because of its assumed relation to the psychophysical function. It should be apparent from Equation 2, however, that a model of the judgmental process is required to interpret Equation 1's parameters. Although most current theories of magnitude scaling consider judgmental processes, the relation of theoretical parameters to the coefficients of regression models is frequently not discussed. In the next section, we introduce models of judgmental and perceptual processes in magnitude scaling that are generalizations of Equations 2 and 3. It is shown that the two regression models discussed (Equations 6 and 11) are closely related to two theoretical models of judgment.

Theory

In this section, we show how the parameters of theoretical models relate to the coefficients of regression models, with an eye toward using the regression coefficients to assess the plausibility of the underlying theory. Although the focus is on judgmental processes in magnitude scaling, the influence of perceptual and memory processes is also considered.

Perceptual and/or Memory Effects

Sequential effects have often been viewed as arising from a perceptual effect of some sort. In particular, the perception, Ψ_t, of a stimulus is assumed to be affected by context. What the relevant context is taken to be marks a point of departure for different theoretical approaches. A general formulation is:

$$\Psi_t = S_t^{\beta} C^{\gamma} \delta_t \tag{13}$$

(DeCarlo, 1989/1990), where C is the context that affects the representation, and δ_t represents error in perception/memory. The parameter γ weights the context: if $\gamma = 0$, then perception depends solely on the current stimulus intensity; as the magnitude of γ increases (either negative or positive), so does the influence of context. Equation 13, which we will refer to simply as the perceptual/memory model, is consistent with the view that perception is relational (Krantz, 1972; Shepard, 1981).

Various alternatives follow from Equation 13 depending on how the relevant context, C, is defined. In Cross's (1973) approach, the context is defined as the ratio (similarity) of the preceding and current stimulus intensities. Substituting (S_{t-1}/S_t) for C in the above equation and rearranging gives:

$$\Psi_t = S_t^{\beta-\gamma} S_{t-1}^{\gamma} \delta_t.$$

It is apparent when written in this form that an assimilative effect (positive γ) will attenuate the exponent obtained for fits of Stevens's power law (the exponent is an estimate of $\beta - \gamma$ and not simply β).

A similar approach is to assume that the relevant context is simply the immediately preceding stimulus, $C = S_{t-1}$, so that Equation 13 becomes:

$$\Psi_t = S_t^{\beta} S_{t-1}^{\gamma} \delta_t. \tag{13a}$$

The only difference between the above model and Cross's model is that Cross's model implies that the coefficient of log S_t obtained for fits of Stevens's power law is attenuated by assimilation (positive values of γ), whereas Equation 13a does not. Because the focus of the present article is on dynamic processes, which are indicated by the coefficients of the *lagged* regressors, Equation 13a is used to simplify the notation. It should be kept in mind, however, that the only change in the models presented below if Cross's (1973) idea is used is that β is replaced by $\beta - \gamma$.

One interpretation of Equation 13a is that the previous stimulus intensity systematically influences the current perception. That is, γ can be interpreted as reflecting a proactive perceptual effect of the previous stimulus, where the sign of γ indicates the direction of the effect. For example, a negative γ indicates a contrast effect. The interpretation in this case is that the current tone, for example, "sounds" softer when preceded by a more intense tone and louder when preceded by a less intense tone. Contrast effects have been found in research concerned with visual illusions and the time error (for references, see Hellström, 1985). A positive γ, on the other hand, indicates an assimilative effect. An example for visual illusions is the well-known "spreading effect" of von Bezold (1886). Hellström (1985) provides examples of assimilative effects in research on the time error.

Another interpretation of Equation 13a is that it models a memory effect of some sort, rather than a perceptual effect. The basic idea in this case is that the current and previous perceptions are "confused" or "assimilate" in memory; Hellström (1985) has noted that ideas of this sort can be traced back to Fechner (1865), Boas (1882), and Köhler (1923). Models of memory effects typically differ with respect to whether the current or previous representations (or both) are considered to be affected. For example, Lockhead and King (1983) have recently presented a model where it is assumed that the current sensation magnitude assimilates in memory toward the previous stimulus magnitude. In this case, both the current, Ψ_t^*, and previous, Ψ_{t-1}^*, representations are affected, because assimilation of Ψ_t implies assimilation of Ψ_{t-1} (this is the approach used in Equation 16 below). Another possibility is that only the sensation magnitude of the previous trial is affected. For example, one interpretation of Luce and Green's model (Equation 7) is that the remembered representation used for comparative judgment, Ψ_{t-1}^*, is systematically different from the original representation, Ψ_{t-1}.

We do not attempt at this point to determine what the "correct" interpretation is, but we simply note that Equation 13a is consistent with basic ideas about perceptual and memory processes. Assuming that $R_t = \alpha \Psi_t \epsilon_t$ (Equation 2 with multiplicative error), where ϵ_t represents judgmental error, substituting Equation 13a and taking logarithms gives:

$$\log R_t = \kappa + \beta \log S_t + \gamma \log S_{t-1} + e_t$$

where $\kappa = \log \alpha$ and $e_t = \log \epsilon_t + \log \delta_t$. The above equation is the relation connecting the observable variables; it shows that multiple regression will provide unbiased estimates of β and γ. If Cross's 1973 idea is used, β in the above is replaced by $\beta - \gamma$, and the model is equivalent to Equation 5, with $\beta_1 = \beta$ and $\beta_2 = \gamma$. In this case, multiple regression will provide unbiased estimates of $\beta - \gamma$ and γ; adding the latter to the former will give an unbiased estimate of β.

Judgmental Processes in Magnitude Scaling

Equation 13a, combined with Equation 2, provides a theoretical basis for Equation 5. It also seems likely, however, that Equation 2 does not correctly or completely represent the way subjects make their judgments. In this section, two models of judgmental processes in magnitude scaling are introduced. The models provide theoretical bases for Equations 6 and 11.

Response Heuristics

The first judgmental model follows from ideas about the role of response heuristics in magnitude scaling experiments. For example, Garner (1953) suggested that (for identification experiments) "there is a tendency on the part of O to repeat what he has just said when he is unsure of his judgment" (p. 379). Similarly, Ward and Lockhead (1971) and Ward (1973, 1979, 1987) have suggested that, in the face of uncertainty, subjects tend to guess in the direction of the previous response. One expression of these ideas is:

$$R_t = \alpha^{1-\lambda}\Psi_t^{1-\lambda}R_{t-1}^{\lambda}\epsilon_t \quad (14)$$

(DeCarlo, 1989/1990) where $0 \leq \lambda \leq 1$ and ϵ_t represents judgmental error (Equation 16 below shows that $\alpha = R_0/\Psi_0$). According to the above, the current response is a geometrically weighted mean of the current sensation magnitude and the immediately preceding response. If $\lambda = 0$, then the above equation reduces to Equation 2 (with multiplicative error), and responses are simply proportional to sensation magnitude. If $\lambda = 1$, then the previous response is repeated, with (possibly) some variation due to ϵ_t (as in cross-modality matching). Thus, λ reflects the tendency to choose a response close to the previous response. Of course, Equation 14 represents only one of a number of possible heuristics that might be used.

We refer to Equation 14 as the *response heuristics* model; the idea has a long history in psychology and related disciplines. For example, in addition to the magnitude scaling references given above, the notion of a response interdependency of one sort or another has been considered by researchers concerned with detection, recognition, probability learning, gambling (for references, see Senders & Sowards, 1952; Treisman & Williams, 1984; Tune, 1964), and the time error (see Hellström, 1985). A similar idea has been discussed in the econometrics literature as the habit-persistence or partial-adjustment model (see Johnston, 1984; Kmenta, 1986).

Assuming that there are both judgmental and perceptual/memory effects in magnitude scaling, in the sense of Equations 14 and 13a, substituting the second equation into the first and taking logarithms gives:

$$\log R_t = (1-\lambda)\beta \log S_t + (1-\lambda)\gamma \log S_{t-1} + \lambda \log R_{t-1} + e_t \quad (15)$$

(the intercept has been dropped), where $e_t = \log \epsilon_t + \log \delta_t$ is assumed to be random.

Equation 15 provides a theoretical basis for Equation 6 (compare the two models to see the equivalence). This relation is important because it explicitly shows how Equation 6's coefficients relate to the theoretical models' parameters (Equations 13a and 14). For example, Equation 15 shows that, because of the assumption that $0 \leq \lambda \leq 1$, a negative β_2 in Equation 6 implies that γ in Equation 15 is negative, since $\beta_2 = (1-\lambda)\gamma$. From Equation 13a, a negative γ indicates a contrast effect, as discussed above. Thus, Equation 15 shows why a negative $\hat{\beta}_2$ for Equation 6 can be interpreted as indicating a contrast effect. Without an explicit presentation of the underlying theory, however, this conclusion is not warranted (see the discussion of Equation 16 below).

Equation 15 also allows one to see whether the coefficients of Equation 6 behave in a manner consistent with the underlying theory. For example, it has previously been suggested (e.g., Garner, 1953; Ward, 1973, 1979; Ward & Lockhead, 1971) that decreasing information about sensation magnitude should increase the tendency of subjects to guess in the direction of the previous response. It follows immediately from Equation 15 that decreasing information should increase the magnitude of the coefficient of $\log R_{t-1}$ (increase λ) and decrease the coefficients of $\log S_t$ and $\log S_{t-1}$, provided that other factors (e.g., β and γ) remain constant. These are, in fact, the same predictions Ward (1979) arrived at using computer simulation of his fuzzy judgment model; the advantage of the present approach is that the predictions follow immediately, without having to resort to computer simulation.

Ward (1979) tested the above prediction using magnitude estimation and cross-modality matching procedures. In both experiments, subjects estimated the distance between two black dots; "information" was (presumably) reduced by decreasing both the presentation time and illumination of the dots. As Ward (1979) noted, and as Table 1 shows, the results for magnitude estimation were as predicted: the estimate of the coefficient of $\log R_{t-1}$ was larger when information was decreased, and the absolute magnitude of the estimate of the coefficient of $\log S_{t-1}$ was smaller. On the other hand, the results for cross-modality matching of duration to distance were not as expected: as Table 1 shows, the estimated coefficients of $\log R_{t-1}$ and $\log S_{t-1}$ were both smaller when information was decreased.

The above equations present the theory that several researchers appear to have in mind when using the terms "assimilation" and "contrast" in more than simply a descriptive manner. Sequential effects arise because the internal representation is affected by context, possibly the previous stimulus, and because of the use of a response heuristic, such as the tendency to choose a response close to the previous response.

The Relativity of Judgment

It has long been recognized that judgment is relative. However, the effect of relative judgment on the sequential structure of psychophysical data has not been fully appreciated. In this section, we present a model of the judgmental process that clarifies the role of reference points in magnitude scaling. The basic idea is that subjects' responses are affected by both immediate context (e.g., the previous sensation and response) and long-term context (e.g., a standard and modulus) as follows:

$$R_t = \Psi_t(R_{t-1}/\Psi_{t-1})^{\lambda}(R_0/\Psi_0)^{1-\lambda}\mu_t \quad (16)$$

(DeCarlo, 1989/1990) where R_0 and Ψ_0 are fixed, or at least relatively constant, references, $0 \leq \lambda \leq 1$, and μ_t represents judgmental error. We refer to Equation 16 as the *relative judgment* model, because it models judgment as being relative to short- and long-term context. The model is similar to earlier judgmental models of visual illusions (e.g., Massaro & Anderson, 1971; Restle, 1971, 1978), in that context is weighted, and is also related to theories concerned with perceptual anchors and response criteria (e.g., Braida, Lim, Berliner, Durlach, Rabinowitz, & Purks, 1984; Gravetter & Lockhead, 1973; Marley & Cook, 1986; Parducci, 1964; Treisman, 1984).

If $\lambda = 0$, then all judgments are made relative to the fixed references R_0 and Ψ_0, and Equation 16 reduces to a stochastic version of Equation 2, with $\alpha = R_0/\Psi_0$. Judgment in this case is often referred to as being "absolute" (the term raises other issues not discussed here). Equation 16 (with $\lambda = 0$) clarifies why many researchers (e.g., Laming, 1984; Parducci, 1968; Postman & Miller, 1945; Stevens, 1956) have stated that "absolute judgment is relative"—absolute judgment is relative to the stable references R_0 and Ψ_0, which represent a *fixed frame of reference.* The frame or context might be a single response-sensation pair (referred to as the modulus and standard by psychophysicists), or might be more than one pair, such as the largest and smallest. The important feature, in terms of sequential effects, is that the frame remains constant over time.

If $\lambda = 1$, then all judgments are made relative to the immediately preceding context (sensation and response), and Equation 16 reduces to the model of Luce and Green (1974a, 1974b), without the assumption of two independent representations. Thus, another view of Equation 16 is that it is a generalization of the response ratio hypothesis of Luce and Green (1974a, 1974b).

It is important to recognize that the difference between the two extremes, $\lambda = 0$ or $\lambda = 1$, is apparent only when the sequential structure of the data is considered. Krantz (1972), for example, focused on mean responses and noted that (for the cross-modality matching paradigm) "If the subject chooses y_i as a match to y_j, which pair x_i, x_j does he use as 'reference levels' to form y_ix_i and y_jx_j? The first matching pair presented? The immediately preceding matching pair? This question loses importance provided that the same match is attained, regardless of what reference match is used." (p. 184). Krantz's statement is correct when interest centers on mean responses. On the other hand, once interest turns to the sequential structure of the data, the reference levels that are used have considerably different implications about the properties of the responses over time. As was shown above, if the reference is solely the previous response-sensation pair, then responses will be autocorrelated, because of the accumulation of judgmental error (the error process of Equation 4 will be nonstationary), whereas this is not the case when fixed references are used. The sequential structure of the data, therefore, provides important information about underlying judgmental processes.

According to the present theory, λ takes on values between zero and one because both short- and long-term context affect judgment. That is, the theory recognizes the influence of two frames of reference, a short-term frame (R_{t-1}/Ψ_{t-1}) and a long-term frame (R_0/Ψ_0), the importance of which only become apparent in a time series analysis. Another view of the model is that it generalizes Equation 2 by replacing α with a parameter that has a fixed (say α_0) and stochastic (say α_{t-1}) component, and more importantly, explicitly defines these components in terms of psychological frames of reference.

Equation 16 has a simple but important relation to Equation 2. The relation can be shown as follows. To simplify the notation, let $\alpha = R_0/\Psi_0$ and rearrange Equation 16 to get:

$$R_t = \alpha\Psi_t(R_{t-1}/\alpha\Psi_{t-1})^\lambda\mu_t\,, \qquad (16a)$$

or more simply

$$R_t = \alpha\Psi_t\epsilon_t, \qquad (16b)$$

where $\epsilon_t = (R_{t-1}/\alpha\Psi_{t-1})^\lambda\mu_t$. It follows from Equation 16b that $\epsilon_t = R_t/\alpha\Psi_t$, which in turn implies that $\epsilon_{t-1} = R_{t-1}/\alpha\Psi_{t-1}$, and substituting this into Equation 16a gives:

$$R_t = \alpha\Psi_t\epsilon_{t-1}{}^\lambda\mu_t. \qquad (16c)$$

The above equation shows that Equation 16 reduces to Equation 2 with an autoregressive error process (the error structure is equivalent to Equation 8 when the model is log-linearized). Thus, the interpretation of autocorrelation in terms of the relative judgment model is that autocorrelation arises because of the relativity of judgment to short- and long-term context. There are, of course, other interpretations of autocorrelation, which in turn suggest different experiments (if any) than those presented below. For example, it could be assumed that λ in Equation 16 is equal to zero (immediate context does not affect judgment), and that autocorrelation arises because R_0 and/or Ψ_0 drift stochastically over time, perhaps because of changes in attention or memory. Luce, Baird, Green, and Smith (1980) have presented a random modulus (R_0) model.

Assuming that there are perceptual/memory effects, in the sense of Equation 13a, substituting into Equation 16c and taking logarithms gives:

$$\log R_t = \beta \log S_t + \gamma \log S_{t-1} + \lambda e_{t-1} + u_t \qquad (17)$$

(the intercept has been dropped), where $e_{t-1} = \log \epsilon_{t-1}$, $u_t = \log \mu_t + \log \delta_t$, and u_t is assumed to be random.[3] Equation 17 provides a theoretical basis for Equation 11 (compare the two models to see the equivalence); it shows that fits of Equation 11 provide direct estimates of the theoretical parameters γ and λ, where γ scales a perceptual/memory effect, and λ indicates the relativity of judgment.

The relative judgment model also has implications for Equation 6. In particular, substituting Equation 13a into Equation 16 (or simply noting the relation between Equations 11 and 12) shows that Equation 17 can be rewritten as:

$$\log R_t = \beta \log S_t + (\gamma - \lambda\beta) \log S_{t-1} + \lambda \log R_{t-1} - \lambda\gamma \log S_{t-2} + u_t \qquad (17a)$$

(the intercept has been dropped). To simplify the situation somewhat, it is assumed that $\log \delta_t$ is negligible relative to $\log \epsilon_t$, that is, most of the variability in responses is assumed to arise from noise in the judgmental process rather than is the internal representation. Only the noise structure is modified if $\log \delta_t$ is considered; the results for the systematic part of the model are unchanged.

Equation 17a is the theoretical counterpart to Equation 6. Its implications, however, are distinctly different from those

[3] We should note that there is more than one way to introduce perceptual/memory effects into the model—it can be assumed that (a) both Ψ_{t-1} and Ψ_t are affected, which suggests a perceptual or memory effect, or (b) only Ψ_{t-1} is affected, which suggests a memory effect. We made the first assumption because the resulting model is simpler. The predictions examined are the same, irrespective of which assumption is made.

of Equation 15. For example, according to Equation 17a, $\hat{\beta}_2$ of Equation 6 is an estimate of $(\gamma - \lambda\beta)$. This means that the finding of a negative $\hat{\beta}_2$ for Equation 6 implies either that γ is negative, or that it is positive, but less than $\lambda\beta$. Thus, Equation 17a, in contrast to Equation 15, shows that fits of Equation 6 do not necessarily lead to the conclusion that γ is negative (contrast); it could be positive (assimilation), as long as $\gamma < \lambda\beta$. Clearly, it is important to be explicit about the underlying theory.

Predictions and Previous Evidence

A basic prediction that follows from the relative judgment model is that any factor that increases the influence of immediate context on judgment will be reflected by an increase in λ. It follows from Equation 17 that $\hat{\rho}$ for Equation 11 should be larger, or equivalently, the estimated coefficient of log R_{t-1} for Equation 6 should be larger. The relevant data have been provided by Ward (1987) and are presented in Table 1. Ward examined two types of instructions in magnitude estimation experiments, known as absolute and ratio magnitude estimation instructions. The crucial difference between these instructions, in our view, is that absolute instructions lead subjects to rely more heavily on long-term context (they are in essence told to ignore the previous response–sensation pair), whereas ratio magnitude estimation instructions explicitly request subjects to use their immediately preceding response and sensation to determine their current response. Comparing the results obtained across the two instructions, it is apparent that the estimate of the coefficient of log R_{t-1} is larger for ratio magnitude estimation than for absolute magnitude estimation (0.544 versus 0.391). Ward's 1987 experiment, therefore, provides between-subjects evidence that supports Equation 16.

A comparison of Equations 15 and 17a shows that the two judgmental models make different predictions about the relation between the coefficients of log R_{t-1} and log S_{t-1} in Equation 6. For example, according to the response heuristics model, an increase in the coefficient of log R_{t-1} will be accompanied by a decrease in the absolute magnitude of the coefficient of log S_{t-1} in Equation 6 (Equation 15 shows that responses are determined to a lesser extent by perception as reliance on the response heuristic increases). According to the relative judgment model, on the other hand, an increase in the coefficient of log R_{t-1} will be accompanied by an increase in the absolute magnitude of the coefficient of log S_{t-1} (because immediate context is weighted more heavily), as long as $\gamma \leq \lambda\beta$ (see Equation 17a, Table 2 suggests that this is the usual case). Table 1 shows that the results of Ward's 1987 manipulation of instructions are in agreement with the prediction of the relative judgment model: an increase in the coefficient of log R_{t-1} was accompanied by an increase in the absolute magnitude of the coefficient of log S_{t-1} (from 0.101 to 0.114), whereas the response heuristics model predicts a decrease.

In sum, Ward's 1987 experiment provides between-subjects evidence that supports the relative judgment model: (a) The coefficient of log R_{t-1} in Equation 6 was larger when the instructions emphasized immediate context over fixed context (ratio versus absolute magnitude estimation instructions), and (b) a positive relation between the coefficients of log R_{t-1} and log S_{t-1} was found. Although these results provide important evidence in favor of Equation 16, an explicit test of the model is needed. In the next section, experiments are presented in which we attempted to directly manipulate the relative weights of short- and long-term context by varying the instructions in a within-subjects design.

Experiment 1: Magnitude Estimation

The relative judgment model implies that it should be possible to manipulate the magnitude of the observed autocorrelation by varying the instructions. The purpose of the first experiment was to test this prediction. Eight subjects participated in two sessions. In one session, subjects were instructed to make all their judgments relative to one response–sensation pair, which they were allowed to choose. This is basically the method of "free" magnitude estimation. In the other session, they were instructed to make all their judgments relative to the immediately preceding response–sensation pair, as in ratio magnitude estimation. The two sets of instructions represent two extremes of Equation 16: $\lambda = 0$ and $\lambda = 1$.

There are two predictions. First, if the relative judgment model is at least partially correct, then the autocorrelation parameter of Equation 11 should be larger when immediate context is emphasized (see Equation 17), assuming of course that subjects attempt to and are able to follow the instructions. In terms of Equation 6, the coefficient of log R_{t-1} should be larger (see Equation 17a). Second, Equation 17a shows that an increase in the coefficient of log R_{t-1} for Equation 6 should be accompanied by an increase in the absolute magnitude of the coefficient of log S_{t-1}. This prediction is contrary to the one that follows from the response heuristics model: Equation 15 shows that an increase in the coefficient of log R_{t-1} should be accompanied by a decrease in the coefficient of log S_{t-1}. The relation between the coefficients of the lagged regressors, therefore, permits a discrimination between the two judgmental models.

Method

Subjects

Eight subjects, undergraduates enrolled in an introductory psychology course at SUNY at Stony Brook, served as subjects; they received course credit for participating in the experiment. Only subjects who claimed to have normal hearing participated.

Apparatus

A General Radio Company oscillator was used to generate 1000-Hz tones. Seventeen tones, ranging from 40 dB to 88 dB (sound pressure), in 3 dB steps, were presented binaurally through Grason Stadler headphones (TDH-39); each presentation was 1 s in duration. The presentation of the stimuli and recording of responses were controlled by an IBM PC. The subject was seated in a sound-attenuating chamber (Industrial Acoustics Company) containing an intercom, a monochrome terminal, and a KAT (Koala Inc.). The KAT is a pad with a surface that maps to the computer terminal; movements of a stylus across the surface of the KAT move an arrow on the terminal.

The stimuli were presented according to sequences generated by the uniform probability generator of SAS (see SAS Institute Inc., 1985a). Only sequences with at least six presentations of each stimulus intensity were used. The autocorrelation function and partial autocorrelation function (see the Appendix) for each sequence were also examined; only sequences with no significant spikes for at least the first five lags were used. A total of six different sequences were used.

Procedure

Each subject participated in two sessions, separated by 1–4 days. The order of instructions was determined by coin tosses. After 4 subjects had received one of the two instructions first, the remaining subjects received the other instruction first, to balance the presentation orders. Each session consisted of 200 trials and took less than 1 hr to complete (including practice trials).

The subjects were first shown how to use the KAT. They were then required to enter five responses to get practice using the KAT. As soon as the fifth response was entered, the instructions for a first practice session appeared on the terminal. The subjects were required to make numerical estimates of eight line lengths, approximately 1.5, 3, 6, 12, 24, 48, 96, and 192 mm in length, presented at least once each, for a total of 12 trials. The instructions were the same as those used in the experiment (appropriate for the condition), with the substitution of the word "length" for "loudness" (see instructions below). Subjects were allowed to proceed at their own pace.

Upon completion of the first practice session, the subjects were given a chance to ask questions and then received the instructions for a second practice session by pressing a button on the KAT. The second practice session consisted of 12 practice trials with 12 of the 17 stimuli used in the experiment. Subjects were told that the purpose was to familiarize them with the range of stimuli used in the experiment (the least and most intense tones were among the 12). In addition, subjects in the fixed reference condition were told to choose their reference loudness and its number during the practice session. Upon completion of the second practice session, the message "You will now be given an opportunity to review the instructions one more time before beginning the experiment. Press the top button to continue." appeared on the terminal. After the button was pressed, the instructions for the experiment appeared and subjects were given one last opportunity to ask questions. They were then able to start the experiment by pressing the KAT's button.

For both conditions, a numerical keypad appeared on the terminal approximately 0.5 s after the offset of each tone. The keypad was a 3 × 4 array consisting of the numbers 0 through 9, a period (for decimals), and a clear entry key (to correct mistakes). A rectangle labeled ENTER was located on the right side of the keypad. Subjects controlled the movement of an arrow located on the keypad (i.e., on the terminal) by moving a stylus across the surface of the KAT. Subjects entered each digit of their chosen number by placing the arrow on a number and then pressing the KAT's button. The final response was entered by moving the arrow to ENTER and pressing the KAT's button. The next trial began approximately 1 s after the response was entered. Subjects quickly became skilled at rapidly entering numbers using the keypad.

The instructions for the *fixed reference* condition were as follows:

> You will next be presented with a series of tones that vary in loudness. Your task is to indicate how loud each tone seems by assigning a number to its loudness. Use only one loudness, any one you like, and its number as a reference point. Try to make all your judgments relative to this reference point. That is, assign your numbers so that the ratio of the current number to the reference number matches the ratio of the current loudness to the reference loudness. You may use any positive numbers you like, including decimals. Do not use zero or negative numbers. If you have any questions, please ask the experimenter now. If not, press the top button to begin.

Note that subjects were allowed to choose their own reference point, which is basically the method of free magnitude estimation.

The instructions for the *prior reference* condition (i.e., ratio magnitude estimation) were as follows:

> You will next be presented with a series of tones that vary in loudness. Your task is to indicate how loud each tone seems by assigning a number to its loudness. Use only the immediately preceding loudness and its number as a reference point. Try to make all your judgments relative to this reference point. That is, assign your numbers so that the ratio of the current number to the previous number matches the ratio of the current loudness to the previous loudness. You may use any positive numbers you like, including decimals. Do not use zero or negative numbers. If you have any questions, please ask the experimenter now. If not, press the top button to begin.

The instructions for the two conditions are identical, except for the specification to use either one loudness and its number as a reference (the fixed reference condition) or the previous loudness and its number as a reference (the prior reference condition).

Results

Mean and variability of responses. The upper panels of Figure 1 present the medians (across subjects) of the mean log responses to each sound pressure level for the fixed reference instructions (left) and prior reference instructions (right). The group data is summarized using medians in lieu of means because the pooled data (across subjects) were (slightly) skewed (the group summaries are virtually unchanged, however, if means are used instead of medians).

For both conditions, the trend is approximately linear. The fixed reference results suggest a break or dip at around 60 dB. Departures from linearity of this sort have frequently been found in magnitude estimation experiments (see Gregson, 1976; Teghtsoonian, 1985).

The lower panels of Figure 1 present, for each stimulus intensity, the medians and interquartile ranges (computed across subjects) of the standard deviations of the log responses. The standard deviations were computed separately for each subject, the medians and interquartile ranges of the standard deviations pooled across subjects were then determined. The figure shows that the variability of log responses tends to decrease with increasing log stimulus intensity, a result also evident in the individual plots (not shown). This result has frequently been found in magnitude scaling experiments (e.g., see Luce, Baird, Green, & Smith, 1980; Marley & Cook, 1986).

Time series analysis of residuals. Figure 2 presents the group autocorrelation (ACF), partial autocorrelation (PACF), and cross-correlation functions (CCF) for the residuals of Equation 4 (see the Appendix for a discussion of these functions). The plots were determined in two steps. First, the ACF, PACF, and CCFs were computed and plotted separately for each subject using PROC ARIMA of SAS (see SAS Institute Inc., 1985b). The medians and interquartile ranges of the correlations were then computed across subjects for each lag

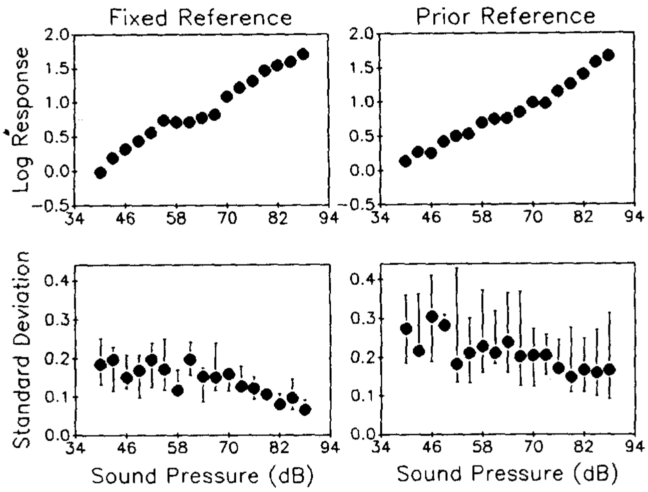

Figure 1. Experiment 1: Magnitude estimation of loudness. The upper panels present, for both instructions, the medians (across subjects) of the mean log responses to each sound pressure level. The lower panels present the medians and interquartile ranges (across subjects) of the standard deviations of the log responses.

and are plotted in Figure 2. The group plots provide a summary of the individual plots.

Both ACFs of Figure 2 show an approximate geometric decay, which is a characteristic of autoregressive processes

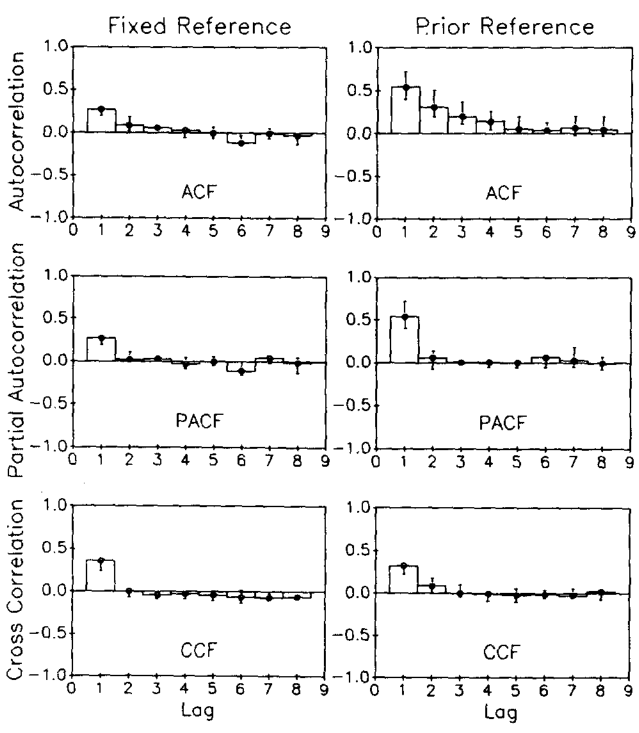

Figure 2. Experiment 1: Magnitude estimation of loudness. The group (medians) autocorrelation functions (ACF), partial autocorrelation functions (PACF), and cross-correlation functions (CCF) are presented for the residuals of Equation 4 (see Appendix for computational details). The functions are plotted separately for each instruction.

(see Box & Jenkins, 1976, and the Appendix). An abrupt drop after the first lag is apparent in both PACFs, which suggests a first-order autoregressive error process.

A comparison of the autocorrelations across the two sets of instructions is of primary interest because the basic prediction of the relative judgment model is that the prior reference instructions will increase the magnitude of the first-order autocorrelation over that obtained for the fixed reference instructions. The ACF and PACF plots of Figure 2 clearly show that the autocorrelation was larger for the prior reference instructions. This result is also evident in the individual analyses, as shown below by the regression analysis.

The cross-correlation functions for both instructions show a positive correlation between $\hat{e}_t$ and log S_{t-1}, which is consistent with the positive relation of Result 1. Although only the cross-correlation for the first lag tends to depart from zero, the CCF plots also suggest a pattern of positive correlation(s) for the first one or two lags, followed by negative correlations for more remote lags. This result may be related to one obtained by Staddon, King, and Lockhead (1980) in an absolute identification experiment.

Deviations from linearity may affect the time series statistics. Although this does not seem likely in this case (because the deviations are small), this possibility was nevertheless assessed by analyzing the residuals obtained by subtracting, separately for each subject, the mean log response for each stimulus intensity from each individual log response. The residuals in this case are estimates of what is referred to as "pure error" in tests of goodness-of-fit (i.e., tests of linearity) for repeat observations (e.g., see Draper & Smith, 1981). No assumptions about the form of the function are made (an ANOVA model is used). Figure 3 presents the ACF, PACF, and CCF obtained (as described above) for the group data using the pooled (pure error) residuals. The plots are virtually identical to those of Figure 2. Thus, the sequential effects do not arise because of departures from linearity.

Figure 4 presents the first-order autocorrelations computed separately for each difference between successive log stimulus intensities (see Appendix for details). The left and right panels of the figure show the results for the fixed and prior reference instructions, respectively. In both cases, an inverted-V pattern appears. The pattern is flatter and slightly higher up (the autocorrelations are larger) for the prior reference instructions than for the fixed reference instructions. The plots show that, in addition to the presence of autocorrelation and cross-correlation, there is a pattern of first-order autocorrelations (the relative judgment model provides a simple interpretation of this pattern, see Discussion below).

Regression analysis. Table 3 presents the estimated coefficients obtained by fitting Equation 4 to each subject's data, the mean estimates (across subjects), the coefficient of determination (R^2), and the *d* statistic of the Durbin-Watson (DW) test (see the Appendix). The upper half of the table presents the results for the fixed reference instructions, the lower half presents the prior reference results. The mean estimate of the coefficient of log S_t is smaller for the prior reference instructions (0.546) than for the fixed reference instructions (0.575). Both estimates are well within the range of those typically found for magnitude estimation of loudness (see Table 1).

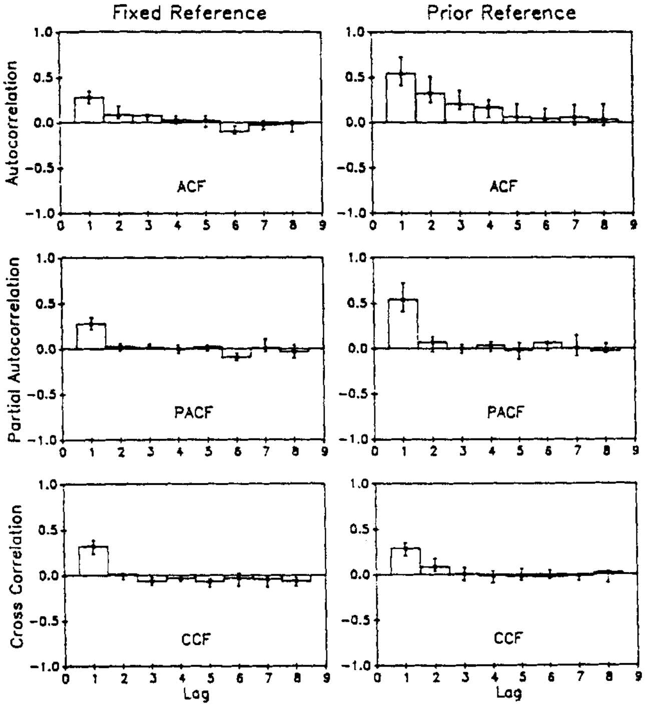

Figure 3. Experiment 1: Magnitude estimation of loudness. The group (medians) autocorrelation, partial autocorrelation, and cross-correlation functions are presented for the residuals from a categorical (ANOVA) analysis (see the Appendix for computational details). The functions are plotted separately for each instruction.

Table 3 also shows that R^2 tends to be smaller for the prior reference instructions than for the fixed reference instructions (see Discussion). The *d* statistic of the DW test indicates the presence of first-order autocorrelated errors (in every case except one), which is consistent with the results shown in Figure 2. The *d* statistic also tends to be smaller for the prior reference instructions, which indicates that the magnitude of the autocorrelation is larger, because $d \approx 2(1 - \rho)$.

Table 4 presents the results for Equation 11. PROC AUTOREG of SAS was used; this procedure gives maximum likelihood estimates of the parameters (see SAS Institute Inc., 1985b, for details). The upper and lower halves of the table present, respectively, the fixed and prior reference results. The results of significance tests on the estimated coefficients of the lagged regressors are also shown (for the individual analyses).

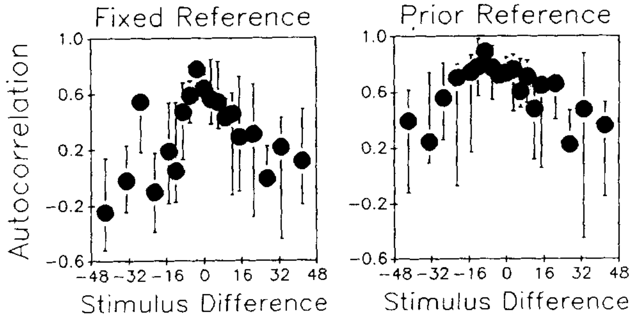

Figure 4. Experiment 1: Magnitude estimation of loudness. The first-order autocorrelations plotted separately for each (nominal) difference between successive log stimulus intensities (see Appendix for computational details).

The mean estimates of the coefficient of log S_t (0.573 fixed, 0.546 prior) for the two instructions are virtually identical to those obtained for Equation 4. The mean estimates of the coefficient of log S_{t-1} (0.084 fixed, 0.096 prior) are positive and similar in magnitude across the two instructions and are also close to those obtained in previous studies (see Table 2). The mean estimate of the autocorrelation parameter, $\hat{\rho}$, is considerably larger for the prior reference instructions (0.557) than for the fixed reference instructions (0.283). For the individual analysis, $\hat{\rho}$ is larger for 7 of the 8 subjects.

Table 5 presents the results for Equation 6. The usual pattern of positive coefficients for log R_{t-1} (assimilation) and negative coefficients for log S_{t-1} (contrast) appears across both instructions. The mean estimate of the coefficient of log R_{t-1} is larger for the prior reference instructions (0.541) than for the fixed reference instructions (0.252); both estimates are close to the $\hat{\rho}$s obtained for Equation 11. The results differ from those of Equation 11 in that the mean estimate of the coefficient of log S_{t-1} is negative and considerably larger for the prior reference instructions (−0.167) than for the fixed reference instructions (−0.065).

Discussion

Magnitude scaling procedures are typically used to obtain information about the form of the psychophysical function and an estimate of its parameters. Another aspect of these procedures, however, is that the temporal structure of the data provides important information about underlying processes. As Figure 1 shows, the mean responses are approximately the same irrespective of what reference points the subjects use. Similarly, Table 3 shows that manipulating the instructions has little if any effect on the exponent of Stevens' power law. On the other hand, the time series analyses show that the sequential structure of the data is greatly affected by the instructions.

The basic goal of Experiment 1 was to determine whether the magnitude of the autocorrelation could be manipulated within subjects by varying the instructions. As Figure 2 and Table 4 show, the observed autocorrelation was considerably larger when subjects were instructed to make their judgments relative to immediate context instead of fixed context. In addition, Figure 3 shows that the sequential effects are unchanged if an arbitrary function (ANOVA) is used in lieu of a linear function. Thus, although autocorrelation in magnitude scaling experiments could arise because any one of a number of (autocorrelated) variables have been omitted from the model (e.g., memory, attention, motivation, etc.), the results support the view that autocorrelation arises in part because of the relativity of judgment to short- and long-term context. Experiment 1 shows that the relative judgment model does more than simply explain autocorrelation, it suggests how to gain experimental control over it as well.

Table 3
Experiment 1: Results for Equation 4

Subject	S_t	R^2	d
	Fixed reference instructions		
S1	.529	.901	1.470*
S2	.665	.876	1.435*
S3	.984	.809	1.245*
S4	.558	.823	1.854
S5	.437	.786	1.538*
S6	.414	.776	1.366*
S7	.448	.828	1.400*
S8	.566	.819	1.609*
M	.575		
	Prior reference instructions		
S1	.415	.845	1.014*
S2	.454	.633	.742*
S3	1.130	.708	1.171*
S4	.522	.817	1.200*
S5	.540	.544	.451*
S6	.310	.466	.482*
S7	.372	.698	.802*
S8	.628	.776	1.503*
M	.546		

Note. d = Durbin-Watson test statistic.
* $p < 0.05$.

Table 3 shows that R^2 for Equation 4 is consistently smaller for the prior reference instructions than for the fixed reference instructions (the result appears for all eight subjects). Ward (1987) also noted a decrease in R^2 (between subjects) across ratio and absolute magnitude estimation instructions. It is simple to show that this decrease is consistent with an increase in autocorrelation. If $e_t = pe_{t-1} + u_t$, it follows from the assumptions about e_t and u_t noted above that

$$\sigma_e^2 = \mathrm{Var}(e_t) = \mathrm{Var}(\rho e_{t-1} + u_t) = \rho^2\mathrm{Var}(e_{t-1}) + \mathrm{Var}(u_t)$$
$$= \rho^2\sigma_e^2 + \sigma_u^2$$

because it follows from the stationarity assumption that $\mathrm{Var}(e_{t-1}) = \mathrm{Var}(e_t) = \sigma_e^2$. Rearranging the last line of the above gives:

$$\sigma_e^2 = \sigma_u^2/(1 - \rho^2)$$

Thus, an increase in ρ leads to an increase in σ_e^2, which in turn leads to a decrease in R^2 for Equation 4 (assuming that σ_u^2 is approximately the same across the two instructions). In words, the accumulation of judgmental error over trials leads to an increase in the variability of responses that is not accounted for by the systematic part of the regression model.

As was discussed above, the two judgmental models (Equations 14 and 16) make different predictions about the behavior of the coefficients of the lagged regressors of Equation 6. In particular, the relative judgment model predicts that an increase in β_3 will be accompanied by an increase in the absolute magnitude of β_2 (at least for values of $\gamma \leq \lambda\beta$, see Equation 17a; Table 2 shows that this appears to be the usual case), whereas the response heuristics model predicts that β_2 will decrease with an increase in β_3. Table 5 shows that an increase in the estimate of the coefficient of log R_{t-1} was accompanied by an increase in the absolute magnitude of the estimate of the coefficient of log S_{t-1} (from a mean of 0.065 to 0.167), in agreement with the relative judgment prediction. Ward's (1987) results also show the predicted increase.

Table 4
Experiment 1: Results for Equation 11

Subject	S_t	S_{t-1}	e_{t-1}	R^2
	Fixed reference instructions			
S1	.524	.053*	.272*	.918
S2	.663	.092*	.267*	.902
S3	.987	.192*	.431*	.872
S4	.553	.084*	.078	.839
S5	.440	.006	.231*	.807
S6	.405	.090*	.376*	.826
S7	.438	.088*	.380*	.870
S8	.569	.069*	.232*	.839
M	.573	.084	.283	
	Prior reference instructions			
S1	.411	.040*	.527*	.904
S2	.441	.102*	.671*	.799
S3	1.117	.245*	.458*	.785
S4	.528	.054*	.416*	.867
S5	.559	.065*	.778*	.823
S6	.314	.081*	.774*	.783
S7	.376	.085*	.625*	.856
S8	.617	.094*	.209*	.795
M	.546	.096	.557	

Note. The coefficient of e_{t-1} was tested for significance using the d statistic computed on the residuals of Equation 5.
* $p < 0.05$.

Although the focus of the present article is on sequential effects, there are other characteristics of magnitude scaling data that have received attention (for a recent list, see Marley & Cook, 1986). For example, Figure 1 shows that the standard deviations of the logarithmically transformed responses tend to be smaller for larger stimulus intensities (this is usually referred to as an "end effect"). Several researchers have considered this result to have implications about underlying processes and have incorporated it into theoretical accounts (e.g., Braida, 1984; Marley & Cook, 1986).

Another result that has received attention in psychophysics is the inverted-V pattern first noted by Jesteadt et al. (1977) and shown in Figure 4 above. The relative judgment model suggests a simple interpretation of this finding: Immediate context is weighted more heavily in judgment when successive stimuli are similar. This suggests modifying the model by, for example, expressing λ as a function of log stimulus differences or by introducing an indicator variable as (for example) follows:

$$R_t = (R_0/\Psi_0)^{1-\lambda D_t}\Psi_t(R_{t-1}/\Psi_{t-1})^{\lambda D_t}\mu_t$$

where $D_t = 1$ if the successive log stimulus difference is less than some value (say 15 dB for loudness estimation) and $D_t = 0$ if it is greater than that value (another possibility is to set

Table 5
Experiment 1: Results for Equation 6

Subject	S_t	S_{t-1}	R_{t-1}	R^2
	Fixed reference instructions			
S1	.525	−.080*	.252*	.917
S2	.667	−.085	.282*	.905
S3	.987	−.167*	.367*	.866
S4	.552	.045	.071	.842
S5	.439	−.100*	.231*	.797
S6	.407	−.040	.316*	.836
S7	.437	−.047	.305*	.877
S8	.568	−.047	.188*	.834
M	.572	−.065	.252	
	Prior reference instructions			
S1	.409	−.167*	.495*	.889
S2	.445	−.164*	.625*	.823
S3	1.132	−.210*	.410*	.790
S4	.530	−.155*	.401*	.855
S5	.576	−.352*	.781*	.828
S6	.342	−.123*	.766*	.840
S7	.390	−.118*	.604*	.862
S8	.617	−.048	.248*	.813
M	.555	−.167	.541	

* $p < 0.05$.

D_t equal to 0 or 1, depending on the log stimulus difference relative to the range of log stimulus intensities used in the experiment). If the residuals from the modified model no longer show the inverted-V pattern, then the model is adequate. However, although modifications such as the above may account for the inverted-V pattern (several other recent theories also offer post-hoc accounts of this result), further research is needed to determine whether or not the modification (i.e., the introduction of additional parameters) is worthwhile (in terms of generating new predictions and experiments).

In sum, the results of Experiment 1 are in agreement with the predictions of the relative judgment model. In the next experiments, we examine the generality and replicability of these results.

Experiment 2: Cross-modality Matching and Magnitude Estimation

The purpose of Experiment 2a was to determine whether the results depend in some way on number usage. Eight subjects were required to produce line-lengths in response to different intensities of a 1000-Hz tone. The experiment was similar in all respects to Experiment 1 (two sets of instructions were used in a within subjects design), with the exception that instead of numbers, responses consisted of line-lengths produced by the subjects.

Experiments 2b and 2c present between-subjects evidence of the effect of varying the instructions. Experiment 2a was a classroom demonstration that served as a pilot for Experiment 1. A class of 52 students made magnitude estimates of the loudness of noise bursts; half the class were given fixed reference instructions, the other half were given prior reference instructions. For Experiment 2c, a class of 18 students made magnitude estimates of the area of circles. Each half of the class received one of the two types of instructions.

Method

Subjects

Cross-modality matching. Eight subjects, undergraduates enrolled in an introductory psychology course at SUNY at Stony Brook, served as subjects; they received course credit for participating in the experiment. Only subjects who claimed to have normal hearing participated.

Loudness and area estimation. The loudness estimation experiment consisted of 52 subjects enrolled in a sensation/perception class at SUNY at Stony Brook taught by L. T. DeCarlo. The area estimation experiment consisted of 18 subjects enrolled in a research methodology class at SUNY at Stony Brook. In both cases, participation in the experiment was voluntary.

Apparatus

Cross-modality matching. The apparatus was the same as that of Experiment 1. Seventeen tones, ranging from 40 to 88 dB in 3 dB steps, were presented binaurally through Grason Stadler headphones; each presentation was 1 s in duration. The order of presentation of the stimuli was determined by the sequences used in Experiment 1.

Loudness estimation. A General Radio Company noise generator was used to generate USASI noise bursts of 2 s duration. Seventeen noise bursts, covering a 48-dB range in 3 dB steps, were presented according to a sequence generated by the uniform probability generator of SAS. The selected sequence had at least four presentations of each stimulus intensity and no significant spikes in the ACF for at least the first five lags. The same sequence was used for both instructions. The noise bursts were presented once every 6 s according to the selected sequence, and were taped using a Revox reel-to-reel tape recorder (model B77 MK II). The Revox recorder was used along with a single speaker to play back the amplified noise bursts to the class. Before the start of class, the most intense noise burst was measured as approximately 104 dB, using a sound meter held about 5 feet directly in front of the speaker.

Area estimation. Nineteen circles were cut out of thin black plastic and were mounted on 282 × 356-mm white cardboard. The diameters of the 19 circles (in millimeters) were approximately 20, 23, 26, 30, 35, 40, 46, 53, 61, 70, 80, 93, 107, 123, 141, 163, 187, 215, and 247. The circles were presented according to a random sequence generated as described above. The same sequence was used for both sets of instructions (except that the order of two of the stimuli differed across the two conditions because of an error made during the stimulus presentations). The circles were presented for approximately 3 s each with an intertrial interval of approximately 6 s.

Procedure

Cross-modality matching. The procedure was identical to that of Experiment 1, with the following exceptions. For the first practice session, subjects were required to enter line lengths in response to the numbers 2, 3, 5, 7, 10, 15, 20, 30, 50, 75, 125, and 200; the 12 numbers were presented in a random order, for a total of 12 trials. For the second practice session (12 trials), subjects assigned line lengths to 12 of the 17 stimulus intensities used in the experiment (the least and most intense were included).

During the experiment, a line length appeared in the middle of the screen approximately 0.5 s after the offset of each tone. The initial length of the line was 1 of 50 values determined by sequences produced by the uniform probability generator of SAS. The lengths covered approximately the full range. Only "initial-length" sequences with no significant autocorrelations for at least the first five lags of the ACF were used. The cross-correlation functions of the stimulus sequences and the initial-length sequences were also examined; only initial-length sequences with no significant cross-correlations with the stimulus sequences for at least the first five lags were used.

The subjects used the KAT to adjust the length of each line. Moving a stylus across the surface of the KAT increased or decreased the line length; movements to the right increased the line length, whereas movements to the left decreased the line length. The smallest line length subjects were able to produce was about 2 mm, the longest was about 203 mm.

The instructions for the fixed reference condition were as follows:

> You will next be presented with a series of tones that vary in loudness. Your task is to indicate how loud each tone seems by assigning a line length to its loudness. Use only one loudness, any one you like, and its line length as a reference point. Try to make all your judgments relative to this reference point. That is, adjust your line lengths so that the ratio of the current length to the reference length matches the ratio of the current loudness to the reference loudness. If you have any questions, please ask the experimenter now. If not, press the top button to begin.

Subjects were allowed to choose their own reference point. The instructions for the prior reference condition were as follows:

> You will next be presented with a series of tones that vary in loudness. Your task is to indicate how loud each tone seems by assigning a line length to its loudness. Use only the immediately preceding loudness and its line length as a reference point. Try to make all your judgments relative to this reference point. That is, adjust your line lengths so that the ratio of the current length to the reference length matches the ratio of the current loudness to the reference loudness. If you have any questions, please ask the experimenter now. If not, press the top button to begin.

The instructions are identical to those of Experiment 1, with the substitution of the words "line length" for "number."

Loudness and area estimation. At the start of class, folders containing response sheets were distributed. The folders were numbered from 1 to 52 for the loudness estimation experiment and from 1 to 18 for the area estimation experiment. Subjects with even numbers participated first, while the other half of the class waited outside the classroom. A coin toss was used to determine which instruction was given first (fixed reference in both cases). The loudness estimation experiment consisted of 120 trials, the area estimation of 118 trials. Each session (including practice trials) took about 30 min to complete.

The subjects were first given a practice session that consisted of 20 trials for the loudness estimation experiment (where each stimulus intensity was presented at least once), and 12 trials for the area estimation experiment (where the largest and smallest circles were included). Subjects were told that the purpose of the practice trials was to familiarize them with the range of stimuli used in the experiment. In addition, subjects in the fixed reference condition (of the loudness estimation experiment) were told to choose their reference loudness and its number during the practice session.

Subjects in the loudness estimation experiment were instructed to make relative judgments of loudness using either one loudness and its number as a reference point (fixed reference instructions), or the immediately preceding loudness and its number as a reference point (prior reference instructions). The instructions for the area estimation experiment were read to the subjects by the experimenter.[4] The instructions were identical to those presented above, except that subjects were asked to judge area. In addition, the middle-sized circle (70-mm diameter) was designated as the standard for the fixed reference instructions (subjects were allowed to choose their own modulus).

Each response folder contained two sheets with the numbers 1 through 30 printed vertically on the front and back of each sheet. Subjects were instructed to slide each sheet up into the folder after they wrote each response, so that responses from previous trials were covered. The folders were collected at the end of each session, and the responses were entered into a spreadsheet.

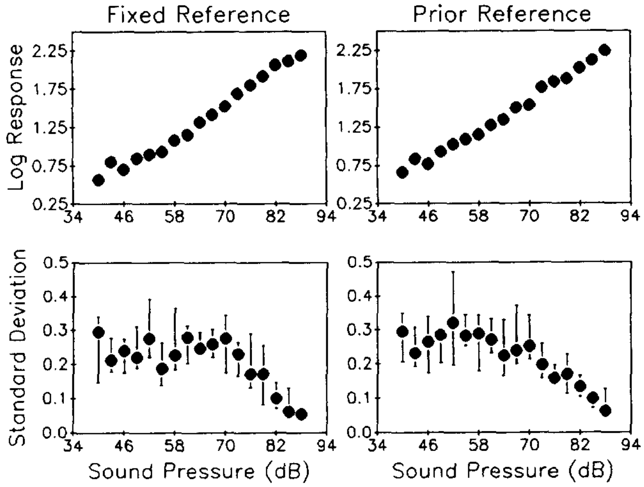

Figure 5. Experiment 2a: Cross-modality matching. The upper panels present, for both instructions, the medians (across subjects) of the mean log responses (line lengths in millimeters) to each sound pressure level. The lower panels present the medians and interquartile ranges (across subjects) of the standard deviations of the log responses.

Results and Discussion

Mean and Variability of Responses

Cross-modality matching. The upper panels of Figure 5 show, separately for each instruction, the medians (across subjects) of the mean log responses to each sound pressure level. The trend is approximately linear. The lower panels present the medians and interquartile ranges (across subjects) of the standard deviations of the log responses. As in Experiment 1, the variability tends to decrease as stimulus intensity increases.

Loudness estimation. The upper panels of Figure 6 present the medians (across subjects) of the mean log responses to each (nominal dB) noise burst for the fixed reference (left) and prior reference instructions (right). The trend in both cases is approximately linear. The lower panels show the medians and interquartile ranges (across subjects) of the standard deviations. The variability of the log responses decreases for higher stimulus intensities. The results replicate those of Experiment 1.

Area estimation. The upper panels of Figure 7 show the medians (across subjects) of the mean log responses to each log area (in square centimeters) for the fixed (left) and prior (right) reference instructions. The trend in both cases is approximately linear. The lower panels present the medians and

[4] We thank Gregory Corona for conducting this experiment.

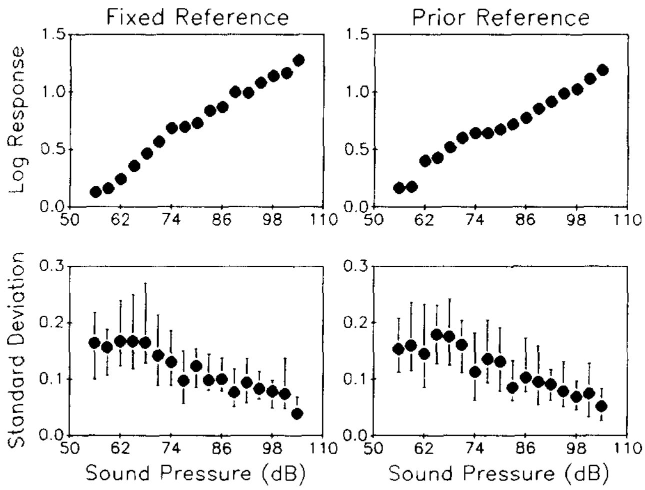

Figure 6. Experiment 2b: Magnitude estimation of loudness. The upper panels present, for both instructions, the medians (across subjects) of the mean log responses to each sound pressure level. The lower panels present the medians and interquartile ranges (across subjects) of the standard deviations of the log responses.

interquartile ranges (across subjects) of the standard deviations of the log responses. There appears to be a decrease in variability at the standard for the fixed reference instructions; this result has frequently been found when visual area is judged (see Baird, 1970; Baird, Green, & Luce, 1980).

Time Series Analysis of Residuals

Cross-modality matching. Figure 8 presents the group autocorrelation, partial autocorrelation, and cross-correlation functions for the residuals of Equation 4, computed as described above. The CCF plots are similar to those obtained in Experiment 1; both show a positive correlation between $\hat{e}_t$ and log S_{t-1}. The critical finding is that the ACF and PACF plots show that the prior reference instructions yielded larger autocorrelation. Thus, manipulation of the instructions increases the autocorrelation, irrespective of whether responses are lines or numbers.

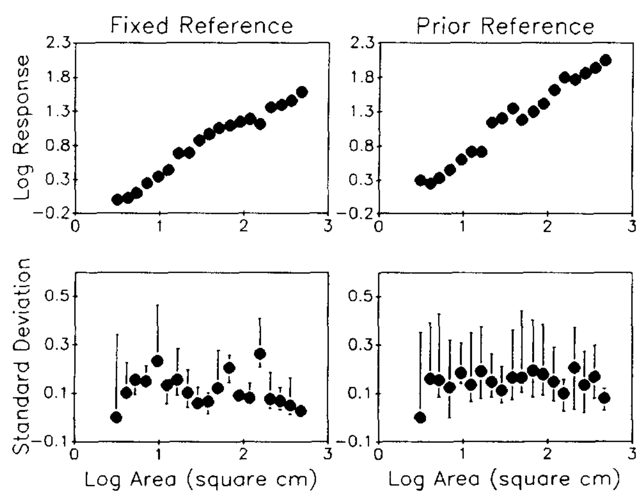

Figure 7. Experiment 2c: Magnitude estimation of area. The upper panels present, for both instructions, the medians (across subjects) of the mean log responses to each log area (square centimeters). The lower panels present the medians and interquartile ranges (across subjects) of the standard deviations of the log responses.

Figure 9 presents the first-order autocorrelations plotted separately for each difference between successive log stimulus magnitudes. The left panel shows the results for the fixed reference instructions, the right panel shows the prior reference results. In both cases, an inverted-V pattern is evident. The results are consistent with those of Experiment 1 and Jesteadt et al. (1977).

Loudness estimation. Figure 10 presents the group autocorrelation, partial autocorrelation, and cross-correlation functions for the residuals of Equation 4. A geometric decay appears in both ACF plots. It is also apparent that the autocorrelation is larger for the prior reference instructions than for the fixed reference instructions. The PACF plots show spikes for the first (and perhaps second) lags. A positive correlation between $\hat{e}_t$ and log S_{t-1} appears in the CCF plots.

Area estimation. Figure 11 presents the ACF, PACF, and CCF for the residuals of Equation 4. The ACF and PACF for the fixed reference instructions show no evidence of autocorrelation. Whether this is a characteristic of area estimation (perhaps because of better visual memory than auditory memory) or whether it resulted from the instruction to use the middle circle as a standard is not known. In contrast, autocorrelation is evident in the ACF and PACF plots for the prior reference instructions. There are no spikes in the CCF for the fixed reference instructions, whereas the CCF for the prior reference instructions shows a small positive spike for the first lag.

Figure 12 presents, for both experiments, the first-order autocorrelations plotted separately for each difference between successive log stimulus magnitudes. The upper panels show the results for loudness estimation, the lower panels show the results for area estimation. An inverted-V pattern is evident for both experiments.

Regression Analysis

Cross-modality matching. Table 6 presents the results for Equation 4. The upper half of the table presents the fixed reference results, the lower half the prior reference results. The mean estimate of the coefficient of log S_t is smaller for the prior reference instructions (0.596) than for the fixed reference instructions (0.700); both coefficients are well within the range of those typically obtained for loudness estimation (see Table 1). The individual fits show that R^2 tends to be smaller for the prior reference instructions than for the fixed reference instructions; the result appears for 7 of the 8 subjects. As was noted above, this result is consistent with larger autocorrelation. The *d* statistic of the DW test indicates the presence of first-order autocorrelation in almost every case, which is consistent with Figure 6. The *d* statistic is also smaller for the prior reference instructions than for the fixed reference instructions (for 7 of the 8 subjects), which, as noted above, indicates larger autocorrelation.

Table 7 presents the results for Equation 11 (see the Appendix). The upper and lower halves of the table present, respec-

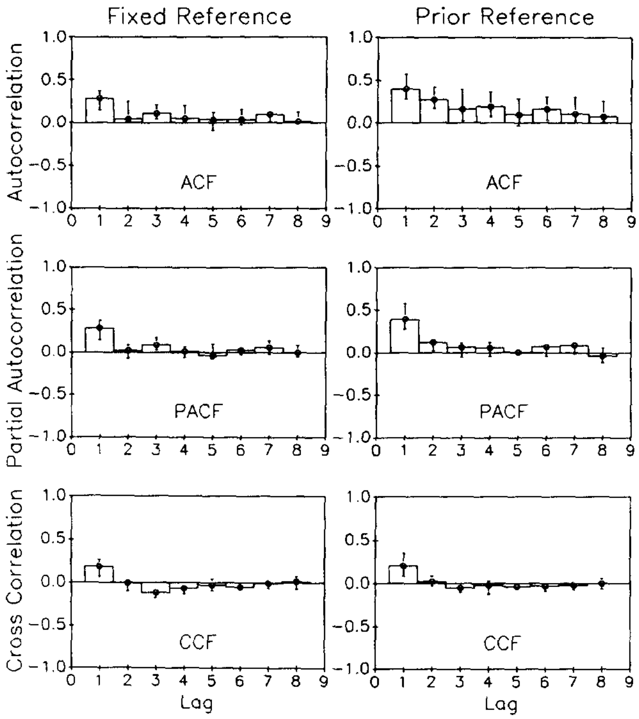

Figure 8. Experiment 2a: Cross-modality matching. The group (medians) autocorrelation functions (ACF), partial autocorrelation functions (PACF), and cross-correlation functions (CCF) are presented for the residuals of Equation 4 (see Appendix for computational details). The functions are plotted separately for each instruction.

tively, the results for the fixed and prior reference instructions. The mean estimates of the coefficient of log S_t (0.695 fixed, 0.591 prior) are virtually identical to those obtained for fits of Equation 4. The mean estimates of the coefficient of log S_{t-1} are similar across the two instructions (0.060 fixed, 0.067 prior) and are close in magnitude to those obtained above and in previous studies (see Table 2). The mean estimate of ρ is considerably larger for the prior reference instructions (0.427) than for the fixed reference instructions (0.277); the result appears for seven of the eight individual cases.

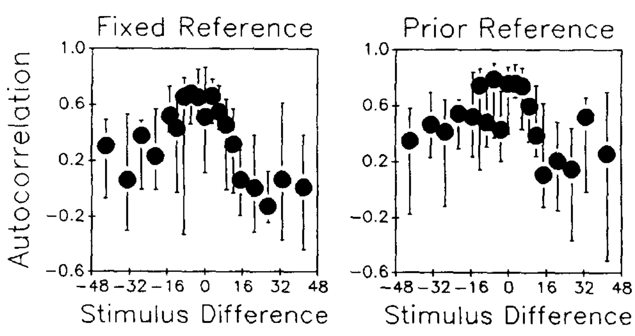

Figure 9. Experiment 2a: Cross-modality matching. The first-order autocorrelations plotted separately for each (nominal) difference between successive log stimulus intensities (see Appendix for computational details).

Table 8 presents the results for Equation 6. The usual pattern of positive and negative coefficients for log R_{t-1} and log S_{t-1}, respectively, appears across both sets of instructions. The mean estimate of the coefficient of log S_{t-1} is larger in absolute magnitude for the prior reference instructions (0.173) than for the fixed reference instructions (0.120). The mean estimate of the coefficient of log R_{t-1} is larger for the prior reference instructions than for the fixed reference instructions (0.417 vs. 0.268); the estimates are close in magnitude to the mean $\hat{\rho}$s of Table 7.

Loudness estimation. Table 9 presents, for both experiments, the mean (across subjects) estimates obtained for fits of Equation 11. For loudness estimation, the mean estimated coefficients of log S_t (0.479 fixed, 0.414 prior) are smaller than those found in Experiments 1 and 2, and the estimated coefficients of log S_{t-1} are small, positive, and virtually identical across the two instructions (0.044 fixed, 0.043 prior). The mean $\hat{\rho}$ is larger for the prior reference instructions (0.421) than for the fixed reference instructions (0.276). Once again, emphasizing immediate context leads to greater autocorrelation.

Area estimation. The mean estimates of the coefficient of log S_t (0.697 and 0.802) are within the range of those typically found for area estimation (see Baird, 1970, Table 3.3, p. 50–51). The mean estimate of the coefficient of log S_{t-1} for the fixed reference instructions is close to zero (−0.006), whereas that for the prior reference instructions is small and positive (0.044). The mean estimate of ρ is considerably larger for the prior reference instructions (0.348) than for the fixed reference instructions (0.035).

Table 9 also presents the results for Equation 6. The usual finding of positive and negative coefficients for log R_{t-1} and log S_{t-1}, respectively, appears in both experiments. The mean estimate of the coefficient of log S_{t-1} is larger in absolute magnitude for the prior reference instructions than for the fixed reference instructions (0.115 vs. 0.082 for loudness estimation, 0.274 vs. 0.030 for area estimation). This agrees with Experiments 1 and 2 above, and with Ward's (1987) results. The mean estimate of the coefficient of log R_{t-1} is larger for the prior reference instructions than for the fixed reference instructions (0.408 vs. 0.265 for loudness estimation, 0.360 vs. 0.041 for area estimation); the estimates are similar in magnitude to the estimates of ρ.

General Discussion

If regression models are to be used for more than merely descriptive purposes, then the relation between their coefficients and the parameters of theoretical models must be made explicit. For example, it was shown above that conclusions about the perceptual/memory process depend on which judgmental process is assumed. In particular, we noted that obtaining a negative value for the coefficient of log S_{t-1} does not automatically lead to the conclusion that the underlying perceptual process is contrastive. Equation 17a, together with Table 2, present another possibility: if judgment is relative, in the sense of Equation 16, then an assimilative influence can appear to be contrastive. Clearly, conclusions about underlying processes must be tempered by a recognition of how

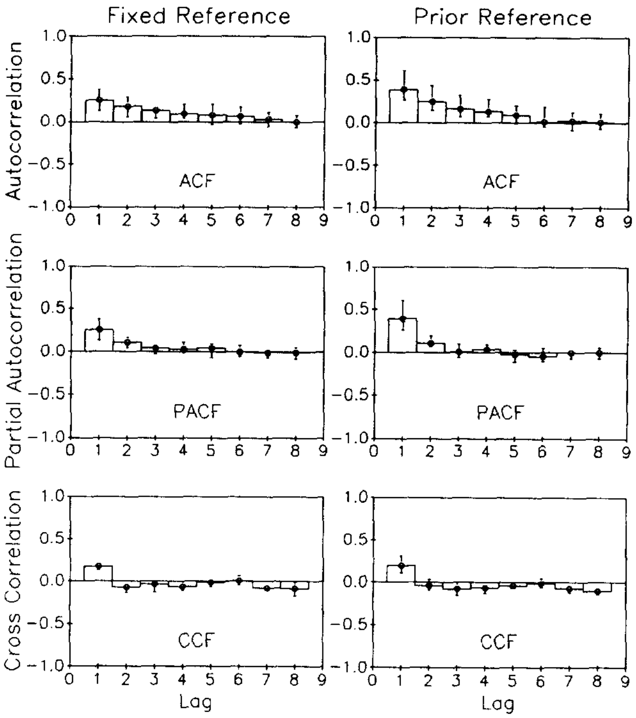

Figure 10. Experiment 2b: Magnitude estimation of loudness. The group (medians) autocorrelation functions (ACF), partial autocorrelation functions (PACF), and cross-correlation functions (CCF) are presented for the residuals of Equation 4 (see Appendix for computational details).

different models, with different interpretations, are interrelated.

We have also stressed the importance of the error term in theoretical and empirical models. Our analysis, for example, shifts attention from the autocorrelation of successive log responses, which has frequently been the focus in prior research, to the autocorrelation of residuals. Several advantages of analyzing the residuals were shown. For example, although Equation 5 accounts for the autocorrelation of log responses, it does not account for the autocorrelation of Equation 4's residuals. Thus, analysis of the residuals reveals that Equation 5 is not complete, whereas analysis of log responses does not. It is also noted in the Appendix that there are advantages to checking for the inverted-V pattern by analyzing the residuals instead of log responses.

We also show that it is important to include the error term in algebraic manipulations of theoretical models. For example, it was shown above that the response ratio hypothesis of Luce and Green (1974a, 1974b; Equation 7) predicts sequential effects if an error term is included in the model, otherwise it does not.

The basic issue addressed in the present article is that, although Equation 6 accounts for the observed autocorrelation, questions remain as to the source of the autocorrelation. According to the response heuristics model (Equation 14), the autocorrelation arises from a response process. In particular, subjects' responses are viewed as being determined by not only sensation magnitude, but also by the use of a response heuristic, such as the tendency to choose a response close to the previous response. The heuristic is typically considered as being used when subjects are uncertain about sensation magnitude. This view motivated Ward's 1979 attempt to test the model by manipulating information. Ward used computer simulation to show that increased reliance on the heuristic should lead to a decrease in the magnitude of the coefficients of log S_t and log S_{t-1} in Equation 6. As was shown above, this prediction follows directly from Equation 14—as reliance on the heuristic increases, responses are determined to a lesser extent by perception, which in turn depends on the current and previous stimulus intensities (see Equation 13a). Ward's 1979 results were somewhat ambiguous: the magnitude estimation experiment supported the predictions, but the cross-modality matching experiment did not. It is also not clear what the "low information" results reveal about judgmental processes in the typical "high information" conditions of magnitude scaling experiments.

According to the relative judgment model (Equation 16), autocorrelation arises because of the influence of different frames of reference on judgment. The basic implication of the model, as shown by Equations 17 and 17a, is that the autocorrelation parameter of Equation 11 or, similarly, the coefficient of log R_{t-1} in Equation 6 can be interpreted as a measure of the relativity of judgment to short- and long-term context. Evidence in favor of this interpretation is provided by Ward's 1987 experiment: the autocorrelation was larger (between subjects) for ratio magnitude estimation instructions than for absolute magnitude estimation instructions. However, rather than simply presenting a post hoc explanation of autocorrelation, we tested the model using a within-subjects

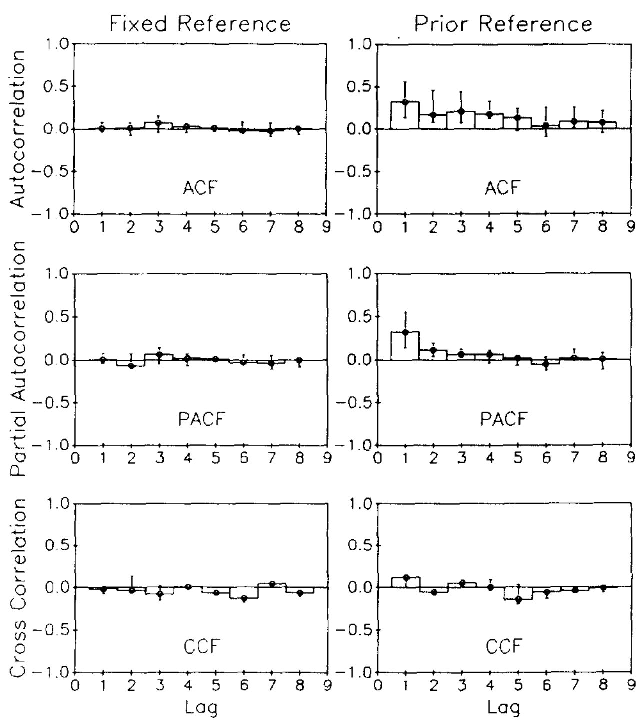

Figure 11. Experiment 2c: Magnitude estimation of area. The group (medians) autocorrelation functions (ACF), partial autocorrelation functions (PACF), and cross-correlation functions (CCF) are presented for the residuals of Equation 4 (see Appendix for computational details).

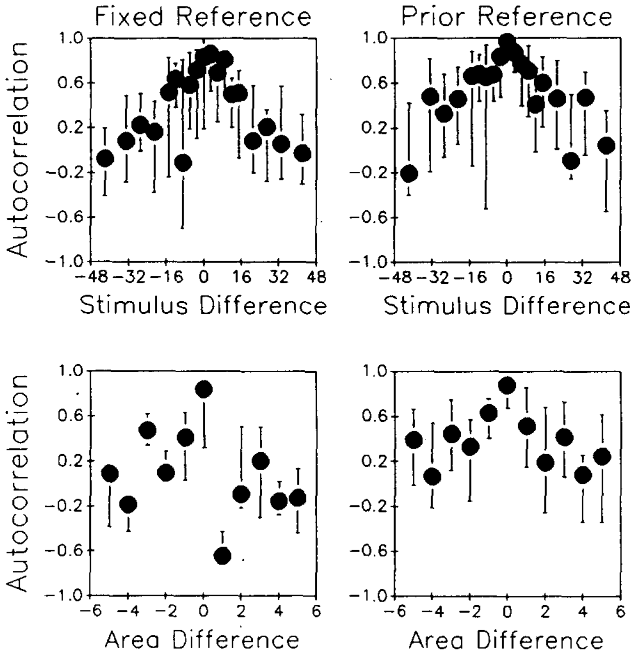

Figure 12. Experiments 2b & c: Magnitude estimation of loudness (upper panels) and area (lower panels). The first-order autocorrelations plotted separately for each (nominal) difference between successive log stimulus intensities (see Appendix for computational details).

Table 7
Experiment 2a: Results for Equation 11

Subject	S_t	S_{t-1}	e_{t-1}	R^2
	Fixed reference instructions			
S1	.659	.039*	.345*	.860
S2	.702	.018	.379*	.892
S3	.565	.063*	.316*	.876
S4	.820	.094*	.285*	.831
S5	.812	.027	.105	.706
S6	.794	.051*	.061	.897
S7	.618	.058*	.454*	.838
S8	.591	.134*	.274*	.752
M	.695	.060	.277	
	Prior reference instructions			
S1	.788	.064*	.468*	.831
S2	.560	.042*	.584*	.893
S3	.390	.018	.140	.711
S4	.642	.136*	.347*	.798
S5	.736	.024	.273*	.747
S6	.585	.063*	.316*	.871
S7	.459	.105*	.592*	.828
S8	.532	.083*	.696*	.795
M	.591	.067	.427	

Note. The coefficient of e_{t-1} was tested for significance using the *d* statistic computed on the residuals of Equation 5.
* $p < 0.05$.

design. If the model is at least partially correct, then it should be possible to manipulate the magnitude of the observed autocorrelation by varying the instructions. All four of the experiments presented above, as well as Ward's 1987 study, show that varying the instructions affects the magnitude of the observed autocorrelation as expected: The autocorrelation

Table 6
Experiment 2a: Results for Equation 4

Subject	S_t	R^2	*d*
	Fixed reference instructions		
S1	.666	.835	1.325*
S2	.670	.858	1.237*
S3	.570	.851	1.396*
S4	.835	.814	1.474*
S5	.810	.700	1.766
S6	.797	.893	1.909
S7	.621	.776	1.019*
S8	.599	.716	1.505*
M	.700		
	Prior reference instructions		
S1	.791	.773	1.073*
S2	.602	.803	.852*
S3	.391	.706	1.703*
S4	.654	.750	1.352*
S5	.727	.713	1.456*
S6	.587	.844	1.387*
S7	.465	.667	.781*
S8	.550	.589	.623*
M	.596		

* $p < 0.05$.

was consistently larger when the instructions emphasized the short-term frame over the long-term frame. The increase in autocorrelation is also consistent with the finding of the present experiments, and Ward (1987), that R^2 for fits of Stevens's power law tends to be smaller for ratio magnitude estimation instructions than for absolute magnitude estimation instructions.

A comparison of Equations 15 and 17a shows that the two judgmental models make different predictions about the behavior of the coefficients of the lagged regressors in Equation 6. The response heuristics model predicts that an increase in the magnitude of the coefficient of log R_{t-1} will be accompanied by a decrease in the absolute magnitude of the coefficient of log S_{t-1}, because responses are determined to a lesser extent by perception as reliance on the heuristic increases. On the other hand, the relative judgment model predicts that an increase in the coefficient of log R_{t-1} will be accompanied by an increase in the magnitude of the coefficient of log S_{t-1}, because of the greater influence of immediate context. All of our results, as well as Ward's (1987) data, support the prediction of the relative judgment model: An increase in the coefficient of log R_{t-1} for Equation 6 was accompanied by an increase in the absolute magnitude of the coefficient of log S_{t-1}.

Although the results support the relative judgment model over the response heuristics model, we do not conclude that the latter model should be dropped; some of the autocorrelation might arise because a heuristic is used. Instead, our goal has been to show that autocorrelation in magnitude scaling arises in large part because of the relativity of judgment to short- and long-term context. Equation 16 is important because it explicitly defines the contexts, clarifies the role of

Table 8
Experiment 2a: Results for Equation 6

Subject	S_t	S_{t-1}	R_{t-1}	R^2
	Fixed reference instructions			
S1	.660	−.186*	.338*	.855
S2	.699	−.259*	.380*	.878
S3	.566	−.105*	.302*	.875
S4	.821	−.131*	.263*	.837
S5	.812	−.063	.113	.703
S6	.794	.019	.037	.898
S7	.625	−.225*	.469*	.844
S8	.589	−.011	.242*	.769
M	.696	−.120	.268	
	Prior reference instructions			
S1	.791	−.298*	.462*	.827
S2	.603	−.298*	.573*	.872
S3	.390	−.040	.144*	.715
S4	.652	−.064	.323*	.812
S5	.735	−.179*	.273*	.735
S6	.586	−.114*	.302*	.867
S7	.470	−.142*	.575*	.845
S8	.552	−.252*	.686*	.813
M	.597	−.173	.417	

* $p < 0.05$.

judgmental error in magnitude scaling, allows us to make specific predictions about the behavior of the coefficients of empirical models, and indicates how to gain experimental control over the time series structure of the data. The model incorporates earlier ideas about absolute and relative judgment and is also in line with recent research suggesting that responses in magnitude scaling experiments represent a compromise between absolute and relative judgment (e.g., Marks, Szczesiul, & Ohlott, 1986; Ward, 1987).

In a recent review of psychophysical scaling, Gescheider (1988) noted that there are "two fundamentally different approaches to research on psychophysical scaling" (p. 183). One "represents the approach of the sensory scientist whose goal is to obtain unbiased scales of sensory magnitude to study sensory processes," the second "represents the approach of the cognitive scientist whose goal is to understand the process of judgment." The present article shows that it is important to understand both cognitive and perceptual processes, irrespective of one's research orientation.

Table 9
Experiments 2b & c: Mean Coefficients for Equations 11 and 6

	Equation 11			Equation 6		
Instructions	S_t	S_{t-1}	e_{t-1}	S_t	S_{t-1}	R_{t-1}
	Magnitude estimation of noise					
Fixed	.479	.043	.276	.479	−.082	.265
Prior	.414	.044	.421	.416	−.115	.408
	Magnitude estimation of area					
Fixed	.697	−.006	.035	.697	−.030	.041
Prior	.802	.044	.348	.800	−.274	.360

References

Baird, J. C. (1970). *Psychophysical analysis of visual space.* Elmsford, NY: Pergamon Press.

Baird, J. C., Green, D. M., & Luce, R. D. (1980). Variability and sequential effects in cross-modality matching of area and loudness. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 277–289.

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. (rev. ed.). San Francisco: Holden-Day.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., & Purks, S. R. (1984). Intensity perception. III. Perceptual anchor model of context coding. *Journal of the Acoustical Society of America, 76,* 722–731.

Cross, D. V. (1973). Sequential dependencies and regression in psychophysical judgments. *Perception & Psychophysics, 14,* 547–552.

DeCarlo, L. T. (1989, April). *Sequential effects in magnitude scaling: An alternative regression model.* Paper presented at the 60th annual meeting of the Eastern Psychological Association, Boston, MA.

DeCarlo, L. T. (1990). Sequential effects in magnitude scaling (Doctoral dissertation, State University of New York at Stony Brook, 1989). *Dissertation Abstracts International, 50,* 3730B.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis.* New York: Wiley.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression, I. *Biometrika, 37,* 409–427.

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression, II. *Biometrika, 38,* 159–178.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology, 46,* 373–380.

Gescheider, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology, 39,* 169–200.

Gravetter, F., & Lockhead, G. R. (1973). Criterial range as a frame of reference for stimulus judgment. *Psychological Review, 80,* 203–216.

Green, D. M., Luce, R. D., & Duncan, J. E. (1977). Variability and sequential effects in magnitude production and estimation of auditory intensity. *Perception & Psychophysics, 22,* 450–456.

Green, D. M., Luce, R. D., & Smith, A. F. (1980). Individual magnitude estimates for various distributions of signal intensity. *Perception & Psychophysics, 27,* 483–488.

Gregson, R. A. M. (1976). Psychophysical discontinuity and pseudosequence effects. *Acta Psychologica, 40,* 431–451.

Hellström, A. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin, 97,* 35–61.

Hibbs, D. A., Jr. (1974). Problems of statistical estimation and causal inference in time series regression models. In H. L. Costner (Ed.), *Sociological methodology* (pp. 252–308). San Francisco: Josey-Bass.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics, 3,* 409–414.

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 92–104.

Johnston, J. (1984). *Econometric methods.* (2nd ed.). New York: McGraw-Hill.

Kmenta, J. (1986). *Elements of econometrics.* (2nd ed.). New York: Macmillan.

Krantz, D. H. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology, 9,* 168–199.

Laming, D. (1984). The relativity of 'absolute' judgements. *British Journal of Mathematical and Statistical Psychology, 37,* 152–183.

Lockhead, G. R., & King, M. C. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 461–473.
Luce, R. D., Baird, J. C., Green, D. M., & Smith, A. F. (1980). Two classes of models for magnitude estimation. *Journal of Mathematical Psychology, 22*, 121–148.
Luce, R. D., & Green, D. M. (1974a). The response ratio hypothesis for magnitude estimation. *Journal of Mathematical Psychology, 11*, 1–14.
Luce, R. D., & Green, D. M. (1974b). Ratios of magnitude estimates. In H. R. Moskowitz (Ed.), *Sensation and measurement*. Dordrecht, Holland: D. Reidel.
Luce, R. D., & Green, D. M. (1978). Two tests of a neural attention hypothesis for auditory psychophysics. *Perception & Psychophysics, 23*, 363–371.
Marks, L. E., Szczesiul, R., & Ohlott, P. (1986). On the cross-modal perception of intensity. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 517–534.
Marley, A. A. J. (1976). A revision of the response ratio hypothesis for magnitude estimation. *Journal of Mathematical Psychology, 14*, 242–244.
Marley, A. A. J., & Cook, V. T. (1986). A limited capacity rehearsal model for psychophysical judgments applied to magnitude estimation. *Journal of Mathematical Psychology, 30*, 339–390.
Massaro, D. W., & Anderson, N. H. (1971). Judgmental model of the Ebbinghaus Illusion. *Journal of Experimental Psychology, 89*, 147–151.
Parducci, A. (1964). Sequential effects in judgment. *Psychological Bulletin, 61*, 163–167.
Parducci, A. (1968). The relativism of absolute judgment. *Scientific American, 219*, 84–90.
Postman, L., & Miller, G. A. (1945). Anchoring of temporal judgments. *American Journal of Psychology, 58*, 43.
Rao, P. (1971). Some notes on misspecification in multiple regressions. *American Statistician, 24*, 37–39.
Restle, F. (1971). Instructions and the magnitude of an illusion: Cognitive factors in the frame of reference. *Perception & Psychophysics, 9*, 31–32.
Restle, F. (1978). Assimilation predicted by adaptation-level theory with variable weights. In N. J. Castellan, Jr. & F. Restle (Eds.), *Cognitive theory* (Vol. 3, pp. 75–91). Hillsdale, NJ: Erlbaum.
SAS Institute Inc. (1985a). *SAS user's guide: Basics* (Version 5 Ed.). Cary, NC: Author.
SAS Institute Inc. (1985b). *SAS/ETS user's guide* (Version 5 Ed.). Cary, NC: Author.
Senders, V. L., & Sowards, A. (1952). Analysis of response sequences in the setting of a psychophysical experiment. *American Journal of Psychology, 65*, 358–374.
Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of 'direct' psychophysical measurement. *Journal of Mathematical Psychology, 24*, 21–57.
Staddon, J. E. R., King, M., & Lockhead, G. R. (1980). On sequential effects in absolute judgment experiments. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 290–301.
Stevens, S. S. (1956). The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology, 69*, 1–24.
Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181.
Stevens, S. S. (1975). *Psychophysics*. New York: Wiley.
Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics, 1*, 439–446.
Teghtsoonian, R. (1985, October). *A skeleton in the psychophysical closet: Sinuosity in magnitude estimation functions. What does it mean?* Paper presented at the meeting of the International Society for Psychophysics, Marseille, France.
Treisman, M. (1984). A theory of criterion setting: An alternative to the attention band and response ratio hypothesis in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: General, 113*, 442–463.
Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91*, 68–111.
Tune, G. S. (1964). Response preferences: A review of some relevant literature. *Psychological Bulletin, 61*, 286–302.
Ward, L. M. (1972). Category judgments of loudness in the absence of an experimenter-induced identification function: Sequential effects and power-function fit. *Journal of Experimental Psychology, 94*, 179–184.
Ward, L. M. (1973). Repeated magnitude estimations with a variable standard: Sequential effects and other properties. *Perception & Psychophysics, 14*, 193–200.
Ward, L. M. (1975). Sequential dependencies and response range in cross-modality matches of duration to loudness. *Perception & Psychophysics, 18*, 217–223.
Ward, L. M. (1979). Stimulus information and sequential dependencies in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 444–459.
Ward, L. M. (1987). Remembrance of sounds past: Memory and psychophysical scaling. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 216–227.
Ward, L. M., & Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology, 84*, 27–34.
Ward, L. M., & Lockhead, G. R. (1971). Response system processes

Appendix

Time Series Analysis of Psychophysical Data

This section reviews the Durbin-Watson test, several correlation functions used in time series analysis, and the method of estimation for Equation 11.

The Durbin-Watson Test

The Durbin-Watson (DW) test (Durbin & Watson, 1950, 1951) is a widely accepted test for autocorrelation that is available in most statistical packages (e.g., SAS, SPSS, BMDP). The test assumes that the error terms follow an AR(1) error process (Equation 8) and determines whether or not ρ is zero. If it is, Equation 8 reduces to $e_t = u_t$, and the errors are uncorrelated. The d statistic of the DW test is computed from the residuals of an ordinary least squares regression as follows:

$$d = \frac{\sum_{t=2}^{T}(\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{T}\hat{e}_t^2} \quad \text{(A1)}$$

where T is the number of trials. If ρ equals zero, then the d statistic is approximately equal to 2. Values of d less than 2 indicate positive autocorrelation, whereas values greater than

2 indicate negative autocorrelation. The DW test is a bounds test; that is, the null hypothesis is rejected if d falls outside lower and upper bounds, d_L and d_U, tabulated by Durbin and Watson and reprinted in many statistical texts. When testing for positive autocorrelation, it is concluded that $\rho = 0$ if $d > d_U$, and that $\rho > 0$ if $d < d_L$; the test is inconclusive if $d_L \leq d \leq d_U$. To test for negatively autocorrelated errors, the test statistic used is $4 - d$.

A significant value of the d statistic means that either (a) the functional form of the model has been misspecified, or the model variables are incorrect (either relevant variables have been omitted, or irrelevant variables have been included), (b) the errors are generated by a first-order autoregressive process, or (c) the errors are generated by some other process. There are three alternatives instead of one because the test is not robust to other violations of the model. For this reason, the DW test is considered to be a test of model misspecification (e.g., see Johnston, 1984).

Correlation Functions

The DW test is relevant for psychophysical research because the evidence discussed in the text clearly indicates the presence of first-order autocorrelation. Higher order autocorrelations, however, should also be examined. We analyze the residuals of Equation 4 using the autocorrelation function (ACF), partial autocorrelation function (PACF), and cross-correlation function (CCF). The ACF is the correlation between residuals separated by lags of 0, 1, 2, and so on. The PACF is similar, except that correlations for intermediate lags are partialed out (see Box & Jenkins, 1976). These plots are used in time series analysis to determine whether or not the process is stationary, to indicate the order of the process, and to determine whether an autoregressive process is appropriate, or whether another process should be considered. For example, it is simple to show for a first-order autoregressive process that the ACF is a geometrically decaying function of the lag. The PACF for an AR(1) process, on the other hand, cuts off abruptly after the first lag (see Box & Jenkins, 1976). Thus, for example, if ρ is positive and the errors are first-order autocorrelated, then the (theoretical) ACF will show a geometric decay over lags, whereas the PACF will show an abrupt drop after the first lag.

The CCF we examine is the cross-correlation of Equation 4's residuals with the current and lagged log stimulus values. This type of plot is used in time series analysis to determine the appropriate transfer function (see Box and Jenkins, 1976). The finding of a positive spike for the first lag in the CCF plots presented above is consistent with the positive relation of Result 1.

Autocorrelation Patterns

We also examine the residuals for the inverted-V pattern discussed in the text. Jesteadt et al.'s (1977) approach was to first fit Equation 4 to each individual's data and to then compute the residual autocorrelations separately for each pair of stimulus intensities. The autocorrelations were then averaged across subjects for each log stimulus intensity difference. The same approach has been used to study the autocorrelation of (log) responses. Luce and Green (1978) pointed out that the response autocorrelations should be computed separately for each stimulus pair because the range of response magnitudes across pairs varies with the range of the corresponding stimulus intensities. However, this approach is only feasible when a large number of observations are available for each subject. The usual means of collecting a large number of observations is to run each subject in several sessions and to then pool the data across sessions. For data sets obtained from single sessions (which usually consist of no more than several hundred trials) not enough observations are available to compute each correlation separately for each pair of stimulus intensities. However, if the autocorrelation of the residuals is analyzed instead of the autocorrelation of responses, then the range problem does not arise, because the range of residual magnitudes across pairs does not vary systematically with the range of stimulus intensities, as response magnitudes do. Thus, the correlations can be computed by fitting Equation 4 to each individual's data and then computing the correlation between $\hat{e}_t$ and $\hat{e}_{t-1}$ for each log stimulus intensity difference, instead of separately for each pair of log stimulus intensities. The advantage of this approach is that it greatly increases the number of residuals used to compute the correlations and makes it possible to check for the inverted-V pattern using much fewer observations per subject.

For the inverted-V plots presented in the text, the successive log stimulus differences were grouped so that the number of observations in each group were roughly equal. For the loudness estimation experiments, the log stimulus differences were grouped into 19 intervals; the abscissae of the plots show the approximate midpoints of the intervals. For the area estimation experiment, the log area differences were grouped into 11 categories. The abscissa of Figure 11 (bottom panel) shows the difference between successive categories. All of the inverted-V figures present the medians and interquartile ranges (across subjects) of the correlations for each log stimulus difference, because the correlations pooled across subjects tended to be slightly skewed. However, the figures are unchanged if the means and standard deviations are plotted instead.

Estimation for Equation 11

Methods of fitting models with autoregressive error processes have been widely investigated in econometrics (e.g., see Kmenta, 1986). The usual method, referred to as *estimated generalized least squares*, is to first estimate ρ using the residuals from a least squares regression (on Equation 4 in this case), transform the dependent and independent variables to remove the autocorrelation, then run a second least squares on the transformed variables. The variables can be transformed as follows:

$$Y_1^* = (1 - \rho^2)^{1/2} Y_1, \quad Y_t^* = Y_t - \rho Y_{t-1},$$
$$X_1^* = (1 - \rho^2)^{1/2} X_1, \quad X_t^* = X_t - \rho X_{t-1}.$$

Note that the transformation preserves the first observation (the above is referred to as the Prais-Winston method in econometrics; see Hibbs, 1974). Ordinary least squares is then performed on the transformed variables Y* and X*, which yields asymptotically efficient estimates. The coefficients presented in the text were obtained using PROC AUTOREG of SAS (for technical details, see SAS Institute Inc., 1985b).

Received October 23, 1989 ■
Accepted May 9, 1990 ■