Journal of Experimental Psychology:
Human Perception and Performance
1983, Vol. 9, No. 3, 461–473

A Memory Model of Sequential Effects in Scaling Tasks

Gregory R. Lockhead and Michael C. King
Duke University

Subjects judge successive stimuli to be overly similar in psychophysical scaling tasks. This is called *assimilation.* They also tend to judge each stimulus as overly different from more previous events. This is called *contrast.* To examine a two-stage linear model of these sequence effects, we asked subjects to judge the relative intensity of successive tones. In support of the model, responses again depended lawfully on prior events. These memory effects occur in a variety of scaling tasks and are consistent with two assumptions: (a) Successive events assimilate in memory, and (b) subjects compare each stimulus to a collection of memories of prior events to generate a response. The trial-by-trial analysis used to test the model also showed that even in magnitude-estimation studies, equal stimulus ratios do not result in equal response ratios, except on average. This article suggests that examinations of trial-by-trial performance might be useful in studying memory and judgment processes.

There are sequential effects in psychophysical scaling data. Responses to successive stimuli are positively correlated in magnitude-estimation, category-judgment, and absolute-identification data. This positive correlation occurs even though successive stimuli are zero correlated. Hence, subjects judge the presented stimulus as overly similar to the judgment of the previous stimulus. This is called *assimilation.* A second sequential effect often exists in the same data. The stimulus is often judged as overly different from stimuli or responses that occurred several trials earlier. This negative correlation between the response to the current stimulus and the value of each of several earlier events is called *contrast.* One or both of these sequential effects can be seen in every reported study in which context effects have been evaluated (cf. Staddon, King, & Lockhead, 1980).

The research for this work was supported by the National Science Foundation. We thank the Department of Experimental Psychology, Oxford University, Oxford, England, and the Psychology Department, Florida Atlantic University, Boca Raton, Florida, where portions of this article were prepared. We also thank Paul Gruenewald, Allan Nash, and Michel Triesman for helpful comments, and Rob Johnson for performing the analysis reported in Table 1.

Michael C. King is now at Bell Telephone Laboratories, Holmdel, New Jersey.

Requests for reprints should be sent to Greg Lockhead, Department of Psychology, Duke University, Durham, North Carolina 27706.

This fact of sequential effects is not consistent with direct-scaling models of psychophysical judgment. An implicit assumption in all such models, which is made explicit in presentations of Stevens' Law, is that "equal physical ratios produce equal subjective ratios" (Stevens, 1957, p. 153). According to this assumption, if two stimuli stand in some physical proportion to one another, and two other stimuli also stand in that same proportion, then the sensed relation between the stimuli in the first set will be the same as that in the second set, excepting random error.

This axiom and sequential effects are incompatible. It cannot be that (a) subjects directly judge and report intensities or relations between intensities and (b) judgments of the stimulus are contingent on prior events.

The most frequent expression of direct scaling views is

$$R_N = kI_N^{\beta}, \tag{1}$$

where R_N is the response on trial N, I is a stimulus measure, and k and β are constants. In practice (Stevens, 1975, p. 14), the logarithm of Equation 1 is taken and the sense magnitude, ψ, is empirically estimated by averaging responses,

$$\log \psi = \log R = \beta \log I + \log k, \tag{2}$$

where R is a random variable representing the response. A primary experimental task for many years has been to determine the values of the constants in Equation 2 for different stimulus domains. Bisection, fractionation, and magnitude production, as well as magnitude estimation (ME), are methods that are frequently used for this purpose. To examine psychophysical judgments in general and to test the power law in particular, it has often been believed appropriate to have observers make proportionality judgments of intensities. Data from hundreds of thousands of such judgments have been reported in the literature.

To produce a response according to Equation 2, the subject must judge the stimulus, multiply that judgment by a remembered constant, and add another constant. Some researchers would object that multiplication, and perhaps addition too, are not essential and would occur naturally. But some calculation must occur because, for just one example, different moduli in ME experiments can result in the same essential scale. Because the identical stimulus is assigned a different response number under a different modulus, the response cannot be a direct judgment in more than one such case.

This is one reason that Cross (1973) suggested that responses in ME experiments depend on proportionality judgments between the current stimulus and the previous stimulus. He suggested that people maintain the same ratio between successive responses as exists in their internal representation between successive stimuli. Jesteadt, Luce, and Green (1977) then suggested that sequential effects are somehow the result of errors introduced by this procedure. This response-ratio-rule (RRR) hypothesis states that

$$R_N = R_{N-1}(aI_N/bI^*_{N-1}). \quad (3)$$

Accordingly, the subject must judge the current stimulus, I_N, divide that magnitude by the remembered value of the prior stimulus, I^*_{N-1}, and multiply the result by the remembered prior response, R_{N-1} (or the subject must perform some analogy to these operations.) If people do all or any of the operations in Equation 3 or 2, and if sequential effects result from those procedures, then reducing the need for such processing should also reduce the sequential effects. The successive-ratios-judgment task (SRJ), which is described in the next section, is such a simplifying task, but the sequence effects were not reduced.

Successive-Ratios-Judgment Task

For the method described here, the subjects are asked to judge the ratio between the current stimulus and the previous stimulus. The explicit instruction is to respond such that R_N reflects the impression of I_N/I_{N-1}. This eliminates any multiplications or additions that may be required by other procedures (e.g., Equation 2). This instruction also removes the requirement to remember the previous response (e.g., Equation 3). Some representation of the previous stimulus, S_{N-1}, must still be remembered for subjects to perform the task reliably. This article examines this important aspect.

This change in the judgment task—from asking for a one-to-one mapping of responses onto stimuli, to asking for a judgment of the ratio between successive stimuli—destroys the usual correlations between stimulus magnitudes and response magnitudes. For the response to be consistent in magnitude estimation (ME) or absolute identification (AI) or absolute judgment (AJ) studies, the same number must be assigned to every occurrence of that stimulus. This is not the case for the ratio-judgment task. Here the same intensity should be responded to with different numbers on different trials. For example, an intensity should be followed by a response number greater than 1 if the previous stimulus had been of lower intensity than it, but a number less than 1 should be assigned if S_{N-1} had been larger than S_N. If performance is perfect, then each intensity will be responded to with as many different numbers as there are intensities in the set. Although there is a 1:1 mapping of stimuli onto overt responses in AJ and ME studies, there is no such mapping here.

The SRJ task thus accomplishes two things. First, compared to the ME procedure, it reduces the formal computational requirements for the subjects; this may simplify their task. Second, it removes the correlation between stimulus intensities and response magnitudes; this may make it simpler for us to determine if prior intensities or prior responses or both determine the sequential effects. Nothing concerning the stimulus presentation method is changed from that for an AJ or ME study. All these methods use the same succession of randomly selected stimuli that vary along one dimension.

Historically, this successive-ratios method has similarities to the ratio method proposed for the study of color by Richardson (1929). In that method, one stimulus served as a standard, and the ratio of an attribute of that stimulus, such as lightness, to the lightness of another stimulus was directly estimated. Although sometimes studied (e.g., Richardson & Ross, 1930; Newhall, 1939), the method is not commonly used. A derivative method, in which people estimate sense differences, is an antecedent of the constant-sum method, which was introduced by Metfessel (1947) as a possible improvement of the fractionation method. Metfessel had people divide 100 points between members of paired stimuli to indicate the perceived relative amounts each stimulus had of some attribute. Following up on that procedure, Baker and Dudek (1955) considered that assigning points to stimuli might require a fair degree of mathematical sophistication on the part of the observer; therefore, they asked subjects instead to indicate directly the ratio between stimuli in a pair. In their studies, as in most research we have encountered that uses the constant-sum method (e.g., Luce & Green, 1974), subjects were not asked to compare any stimulus with itself. Perhaps this is because of the assumption that, excepting random error, this ratio is 1. Some researchers, have included sitmulus repetitions (e.g., Richards, 1974) but have usually averaged their data over trials such that the presented results are incomplete for the purposes here.

An SRJ Experiment

The primary purpose of this experiment was to determine what performance is like, and whether there are memory effects when people are asked to make relative judgments of the intensities of successively presented tones.

Method

Subjects. The subjects were a young man and woman who had not served in a SRJ study before.

Stimuli. The stimuli were 30, 1000-Hz sinusoids of ½-sec duration that differed in 1-dB steps from 51 to 80 dB SPL.

Procedure. The individually tested subjects sat in a sound-attenuating (IAC) chamber. Stimuli were presented binaurally through headphones. Responses were made on a keyboard composed of a 3 × 3 matrix of numerals, a zero key, a decimal-point key, and an "enter" key. Each response was displayed on a video screen in front of the subject and was followed by a display of the feedback as soon as the subject pressed the "enter" key on the keyboard. The screen then erased and the next randomly selected tone was presented 2 sec after the enter key had been pressed. The instructions to the subjects were as follows:

> You will be presented with many tones varying in loudness and are asked to respond to each tone by typing a number representing the ratio of the present loudness to the loudness of the just previous tone. For example, if the present tone is twice as loud as the prior one, type 2 CR (carriage return). If it is half as loud, type .5 CR, if it is equally loud, type 1.0 CR. Please use decimals accurately (i.e., to as many places as you wish). The first two tones are each the middle stimulus in the loudness series—respond to the second by typing 1 CR. You will be given feedback immediately after each response. The feedback represents the ratio of the subjective intensities as measured by the average subject. The range of ratios that can occur is from .10 to 9.18. Please respond quickly but accurately. There will be 100 trials per block and 4 blocks in this session. There is a total of five sessions in the experiment. Any questions?

To provide feedback, each physical intensity was raised to the .67 power, the exponent most commonly reported as the best fit to Stevens' law for middle-frequency-range loudnesses, and the ratio between these numbers for successive stimuli served as the feedback to the observers. The feedback numbers ranged from .10 to 9.18. Each subject provided 396 responses on each of 5 daily sessions of approximately 40-min. duration.

Results

No consistent differences between the subjects were detected, and their data have been combined for the following analyses. Any statement concerning the combined data is also true for both individuals' sets of data.

Average performance. Figure 1 shows the mean ratio judgment given to each stimulus ratio. A total of 161 different stimulus ratios occurred in the study. A data point in the figure represents the mean of 2 to 122 judgments. Beyond the reason of random selection, this variability in the number of observations per point occurs because the distribution of stimulus ratios is not uniform whenever physical intensities are presented randomly. For example, stimulus repetition

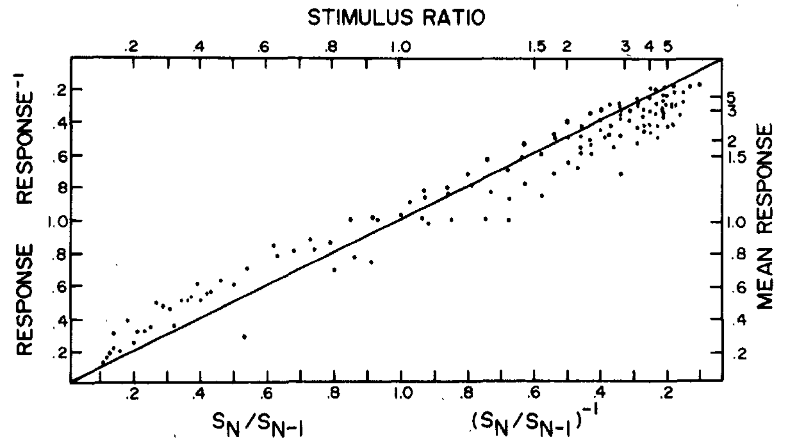

Figure 1. The mean response to each stimulus ratio (S_N/S_{N-1}).

(ratio = 1.0) is expected on 1/30 of the trials because there are 30 different stimuli being randomly selected with replacement, but the largest stimulus ratio can only occur when the 80-dB tone follows the 51-dB tone, an expected 1/900 of the trials.

The solid line in Figure 1 represents the feedback provided after each response. If the subjects performed the task perfectly, then all of the data points should fall on that line. Although this tends to be the case, a consistent error in the responses can be seen in Figure 1. Stimulus ratios less than 1 are overestimated and ratios greater than 1 are underestimated.

The analysis shown in Figure 1 confounds stimulus ratios with stimulus intensities. Because responses depend on S_{N-1} as well as on S_N, the same stimulus ratio could often occur whether the current stimulus was quiet or loud. To learn if performance is related to stimulus intensity, Figure 2 shows the mean response as a function of S_N, with S_{N-1} as a parameter. The just prior stimuli have been categorized into three groups of 10 tones (quiet, medium, and loud) for convenience of presentation. Responses tend to increase monotonically and about linearly with physical intensity.

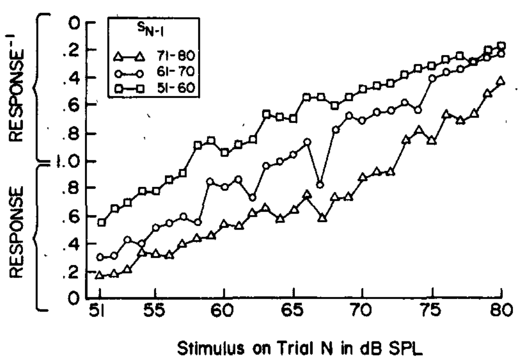

Figure 2. The mean response to S_N/S_{N-1} following each stimulus intensity (S_N) when the stimulus on the prior trial (S_{N-1}) was a low (squares), medium (circles), or high (triangles) intensity.

Stimulus and response repetitions. Ratios and intensities are confounded in Figures 1 and 2. To separate these, we considered only trials in which $S_{N-1} = S_N$. This is the unique case in which, no matter what the psychophysical transformation between stimuli and responses might be, the appropriate response is 1 if there are no memory effects. The left panel of Figure 3 shows the mean response as a function of stimulus intensity when $S_{N-1} = S_N$. This stimulus ratio of 1 was underestimated for low-intensity stimuli and overestimated for high-intensity tones. The mean responses that were averaged to produce Figure 3 ranged from .3 (the 52-dB tone) to 2.35 (the 74-dB tone), a factor of nearly 8. Only two of these average responses were less than 1 when the repeated stimulus was 64-dB loud or louder, and only one such response was greater than 1 when the repeated stimulus was less than 64 dB. The correlation between responses and intensities (in decibels) for repeating tones is .60. Although the

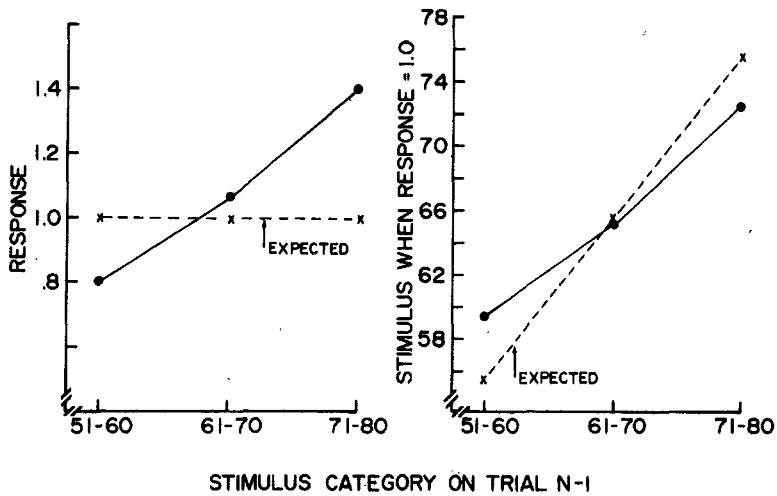

Figure 3. The mean response on trial N (dots, left panel) when successive stimuli (S_N and S_{N-1}) were physically identical, and the mean stimulus on trial N (dots, right panel) when successive stimuli were identified as being identical, as a function of the stimulus intensity on trial $N-1$. (The data have been averaged into three groups of 10 prior stimuli, low, medium, and high intensity. The Xs show the responses expected if there are no sequence effects.)

average response across all repeating stimuli is close to the expected value of unity (see Figure 1), 1 was seldom given on any particular stimulus repetition trial.

The right panel of Figure 3 shows the average intensity judged equal to S_{N-1} as a function of the intensity of S_{N-1}. A quiet S_N was judged as identical (response = 1) to the previous stimulus only if S_N was more intense than S_{N-1}. The opposite result occurred when S_N was loud. Now S_N had to be quieter than S_{N-1} for the two tones to be judged identical. These results allow the suggestion that subjects first encode (or identify) each stimulus and then generate a ratio by comparing the coded stimulus with the memory of the previous stimulus.

A Model

This model of stimulus identification is based on assimilation, which can be defined as stimuli being identified as overly similar to prior responses, prior stimuli, or prior feedbacks. We first examine two assumptions and then discuss the relevance of the model to the SRJ data.

The first assumption is that each presented stimulus, S_N, assimilates toward the memory of the just previous stimulus, M_{N-1}. This may be a passive process such that the perception of any particular stimulus depends on previous events, or it may be an active process due to categorization. No specific process or mechanism is assumed. There are too many possibilities. One candidate mechanism might be that each stimulus sensitizes some channel such that the subsequent stimulus travels that route more readily than it does an unbiased route. A different candidate might be a response or attention system. Perhaps attention moves from the prior encoding toward the current stimulus but there is an inertia such that it moves only part of the total distance, with the stimulus coded or localized as that stopping point.

Whatever process or mechanism is involved, one immediate implication of assuming that successive stimuli assimilate in memory is that M_{N-1} is different from S_{N-1} in a predictable way. The perceived or encoded S_{N-1}, M_{N-1}, had assimilated toward M_{N-2}. That earlier memory, M_{N-2}, is partly a result of assimilation of S_{N-2} toward M_{N-3}, and the memory of M_{N-3} is a result of S_{N-3} assimilating toward M_{N-4}, and so on. In this manner, stimulus events that took place many trials earlier may become reflected in the current response. The subject identifies the assimilated S_N, not the uncontaminated S_N that had been presented.

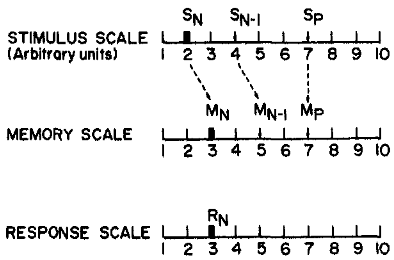

Figure 4. The stimulus, S_N, is assumed to assimilate toward the memory of the prior stimulus, M_{N-1}, and is thus overestimated in the response, R_N.

The second assumption is that the subject decides upon a response by comparing this encoded event to memories of prior events. These include R_{N-1}, the prior feedback if available, and the mean of events on trials $N-2$ through $N-6$ (King, 1980).

Some features of the model are outlined in Figure 4 for a 10-stimulus identification task. The top portion of Figure 4 represents the physical stimulus scale in which stimuli increase in intensity in arbitrary units. The mean of stimuli S_{N-2} through S_{N-6}, called the *stimulus pool,* was 7 units. For this example, S_{N-2} was also 7 units. The just prior stimulus, S_{N-1}, was 4 units, and the stimulus to be judged, S_N, was 2 units.

The bottom portion of the figure represents the response scale. S_N has been overestimated. The reason for this error is diagrammed in the center of Figure 2 and is described here. The memory pool, M_P, has the same value as the mean of the stimulus pool, S_P, for this example. Because M_P depends on events previous to S_P, these will not generally be identical. When S_{N-1} was presented, it assimilated toward M_{N-2} and became represented as M_{N-1}. The eventual S_N assimilated toward this M_{N-1} and thus was overestimated in the response.

Ignoring scaling constants for simplicity, this can be stated as

$$R_N = S_N + a(M_{N-1} - S_N), \qquad (4)$$

where S_N is the stimulus intensity on trial N, M_{N-1} is the remembered value of the stimulus on the previous trial, and a is a positive constant. Equation 4 can only account for assimilation, and this is what is considered in the following description. An account of contrast is suggested in the Assimilation May Produce Contrast section.

The Model and the SRJ Data

To generate an SRJ response, the encoded S_N (see Figure 4) is compared by the subject with M_{N-1}. Consider when a quiet tone is repeated. On average, the stimulus pool is greater than the first such tone, so that the stimulus is assimilated upward. As a result, the subsequent tone is "quieter" than its identical predecessor and is assigned a ratio response of less than 1 (see Figure 3).

This result is not just a scale shift, as might be predicted by sensory adaptation or by a sinking-memory-trace theory in which all memories decrease or increase in intensity. This is seen by noting in Figure 3 that the opposite effect occurs when identical stimuli are loud. Now S_{N-1} is remembered as quieter, not louder, and the response is larger than 1, not less than 1. If there were only a scale shift, the reported relation between S_N and S_{N-1} would not be a function of their intensities.

Responses and Prior Events

Performance with repeating stimuli was next examined when the stimulus just before these two, S_{N-2}, had been loud or quiet. The most appropriate analysis for this purpose is to divide the data of Figure 3 according to the intensities of prior stimuli. However, these data are already only $\frac{1}{30}$th of the total, and there are not enough observations to subdivide them much further. For the present analysis, stimuli were collapsed into three categories of 10 stimuli each, 51–60, 61–70, and 71–80 dB (low, intermediate, and high, respectively). If the model is appropriate, this averaging should not affect the form of the results. Whether identical stimuli or approximately identical stimuli repeat, average responses should be small when both stimuli are quiet and large when S_N and S_{N-1} are both loud. The positive slopes of the functions in Figures 5, 6, and 7 show that this general result did occur.

Figure 5 shows the mean response when a stimulus category repeats; the parameter is whether the stimulus on trial $N-2$ was more or less intense than the series mean. There is an essentially constant effect on responses of whether the stimulus on trial $N - 2$ was greater or less than the mean log intensity. When S_{N-2} was loud, responses to stimuli that belong to the same physical category were regularly smaller than when S_{N-2} was quiet.

To examine the importance of earlier events, we then considered the contribution of the stimulus pool. Figure 6 presents the same analysis as that in Figure 5, but now with the mean of S_{N-3} through S_{N-6} as the parameter. S_{N-2} was ignored in this analysis. When this stimulus pool, S_P, was greater than the overall mean (65.5 dB), responses were smaller than when S_P was less than the mean.

Figure 7 shows the combined effects of S_{N-2} and S_P when S_{N-1} and S_N are in the same category. These combined sequential effects on the ratio judgment are largest of all.

Not all of the memory effect in this experiment is captured by S_{N-2} and S_P in Figures 5 through 7. Consider the result in Figure 7 when S_P and S_{N-2} were both loud. S_N was still judged louder than S_{N-1} when these stimuli were high and quieter when they were low. If S_{N-2} and S_P were the only contributors to the memory, then responses here should be about 1. This suggests at least two possibilities: (a) The measure of context used here is inadequate. Perhaps the combination rule should be nonlinear or some other linear combination than that used in the present analysis. (b) Events further back in the sequence than six trials are important to the judgment. Some long-term overall average of all stimulus memories, perhaps an overall mean or adaptation level (Helson, 1948), could account for the effects remaining in Figure 7 after the context considered here is accounted for.

There is a strong suggestion in the literature that long-term memory effects are indeed important to judgment. In an AJ experiment, subjects provided 500 responses to a set of loudnesses on one day. On the next

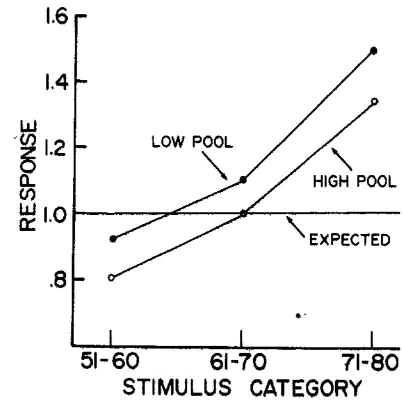

Figure 6. The mean response when successive stimuli (S_N and S_{N-1}) belonged to the same category and the stimulus pool was less than (low) or greater than (high) the average pool value. (The value of S_{N-2} was not considered for this analysis.)

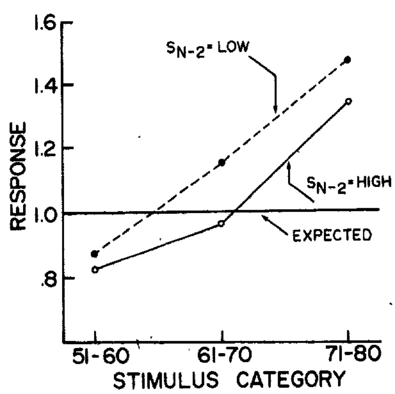

Figure 5. The mean response when successive stimuli (S_N and S_{N-1}) belonged to the same category (low, middle, or high intensity) and the just previous stimulus intensity (S_{N-2}) was less than (filled circles) or greater than (open circles) the average intensity of 65.5 dB. (The horizontal line shows the expected mean responses in the absence of context effects.)

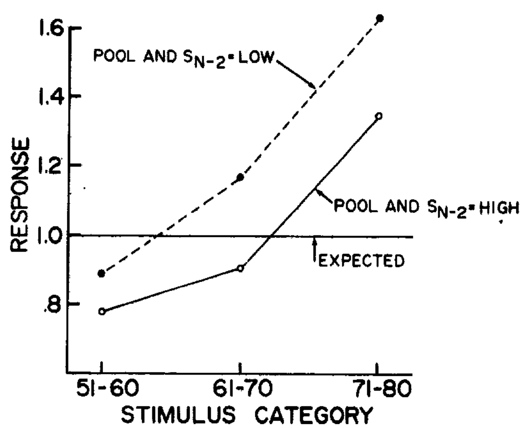

Figure 7. The mean response when S_N and S_{N-1} belonged to the same category and when S_{N-2} and the stimulus pool were both lower (filled circles) or higher (open circles) than 65.5 dB.

day, unknown to the subjects, the entire stimulus scale was shifted up or down by 5 dB and the experiment was repeated. This produced a substantial constant error, even though feedback was given on every trial:

> The effect of the scale responded to on the previous day was surprisingly tenacious and generally still apparent in the last 200 trials (following 300 trials of absolute judgments with feedback) of each shifted day. (Ward & Lockhead, 1970, p. 32)

In that study, stimuli must have been compared with memories from 24 hours previously. In light of that memory effect, perhaps it is not so remarkable to suggest in the current SRJ study that events that occurred more than six trials earlier contributed to performance.

Successive-Ratio Judgment as an Absolute-Judgment or Magnitude-Estimation Task

We have suggested that SRJ data are the result of categorizing or identifying the current stimulus and then dividing that judgment by M_{N-1}. This first step, that of determining "what this thing is," may also be the normal procedure in AJ and ME tasks as well as in everyday object identifications. To evaluate this suggestion of a common process it is necessary to have the data from the various experiments in a common form. For this purpose, the SRJs were transformed into the equivalent of AJ or ME judgments.

In an SRJ study, by instruction, the response on any given trial, N, is $R_N = I_N/I_{N-1}$, where I is stimulus intensity. In general, if subjects follow instructions,

$$R_N = S_N/M_{N-1}, \tag{5}$$

where R_N denotes the random variable representing the response on trial N, S_N is the random variable denoting the internationalization of the signal on trial N, and M_{N-1} is the random variable denoting the memory of the signal that occurred on trial $N-1$.

In a response-ratio-rule (RRR) task (Cross, 1973), subjects are instructed to maintain the same ratio between successive responses as exists between successive stimuli. In general,

$$R_N = (S_N/M_{N-1})R_{N-1}. \tag{6}$$

This differs from Equation 5 only by multiplication with R_{N-1}.

In an ME task, the essential instruction is that $R_N = I_N$, and the general theoretical expression is Equation 1. If there are no sequential effects due to the stimulus or response on trial $N-1$, then multiplying I_N by R_{N-1}/M_{N-1} introduces no average effect, and ME data and RRR data are logically identical (Luce & Green, 1974). By the same assumption, ME and SRJ data are identical if the right side of Equation 5 is multiplied by M_{N-1}. Therefore, multiplying each SRJ (Equation 5) in the present experiment by the appropriate R_{N-1} essentially produces RRR or ME data. SRJs multiplied by R_{N-1}, RRR judgments, MEs, and AJs are all formally equivalent except for scale transformations.

An immediate implication of these results is that it is appropriate to learn if sequential effects in the various tasks are also equivalent in form. For this purpose, each ratio judgment in the SRJ study was multiplied by its respective S_{N-1} intensity raised to the .67 power (cf. King & Lockhead, 1981) to convert the SRJ data to estimated ME data. These stimulus values were used, rather than responses, to derive estimates of the identifications so that any response errors are constrained to the trial in which an error occurred. Any sequential effects found to extend beyond the prior trial thus cannot be due to an induced response propagation. The algorithm used to adjust the data was

$$ME_{\text{est}_N} = R_N \cdot S_{N-1}{}^{.67}, \tag{7}$$

where ME_{est_N} is the estimated response that would have been given to S_N under ME instructions, R_N is the response actually given in the SRJ task, and S_{N-1} is the stimulus intensity on the prior trial. This translates SRJ data such that they are formally analogous to AJ and ME data.

Application of Equation 7 to the SRJ data provides a string of number pairs. Each pair is a physical stimulus intensity, S_N, and an estimated response magnitude, ME_{est_N}. The ordering of these pairs in the string is the order in which the stimuli and responses had occurred in the SRJ experiment. Sequential effects in this string were examined by calculating what contingencies exist between each ME_{est_N} and previous events (see Staddon et al., 1980). The results, with prior stimuli collapsed into groups of five for presentation convenience, are shown in Figure 8.

Consider the effect of the previous stimulus intensity. ME_{est} tends toward the intensity on trial $N-1$ (lag 1 in Figure 8). When S_{N-1} was a quiet tone (open squares), the derived response to all current stimuli tended to be low. With increases in the intensity of S_{N-1}, ME_{est} also increased. This is assimilation.

Figure 8 also shows that the derived response tends to be large if the stimulus on trial $N-2$ was small, and tends to be small if S_{N-2} was large. This is contrast. There appears to be some contrast from two to about six trials back. The analysis was continued for eight trials back, but there is no indication of any consistent trend beyond that shown.

For comparison purposes, Figure 9 reproduces the sequential effects reported when the identical 30 stimuli studied here were used in an ME task with feedback provided and with different subjects (King & Lockhead, 1981, Figure 2). Those subjects were instructed to directly judge each stimulus intensity rather than to judge the ratio between successive intensities. The results in Figures 8 and 9 are essentially identical. Assimilation is followed by enduring contrast in both data sets. It is thus appropriate to consider that the processes producing these various data are also similar.

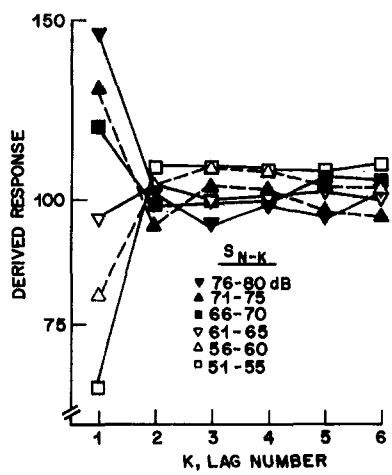

Figure 8. The mean derived response to all stimuli as a function of the physical intensity of the stimulus that occurred from one to six (k) trials earlier in the sequence. (Prior stimuli have been collapsed into six groups of five adjacent intensities.)

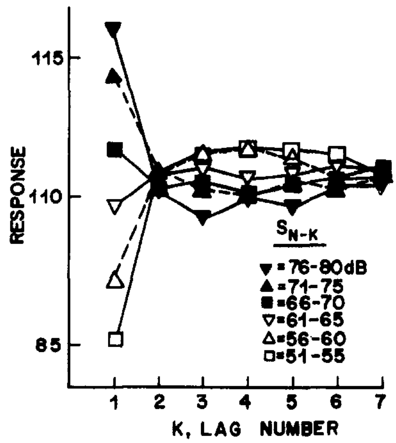

Figure 9. The actual response to all stimuli as a function of the physical intensity that occurred from one to seven (k) trials earlier in the sequence in a magnitude-estimation-with-feedback experiment using the same 30 stimuli as the successive-ratios-judgment study. (From "On Memory Effects in Magnitude Estimation Experiments" by M. C. King and G. R. Lockhead, *Perception & Psychophysics*, 1981, *30*, 599-603. Copyright 1982 by Psychonomic Society, Inc. Reprinted by permission.)

Assimilation May Produce Contrast

Equation 4 describes assimilation but does not account for the contrast seen in Figures 8 and 9 and in the literature. Contrast may result because subjects attempt to keep track of labels for the previous stimuli to maintain a reliable response scale. For example, consider an AJ-with-feedback study using intensities 1–10. Suppose Stimulus 9 was presented and responded to correctly, and then Stimulus 5 was presented. Because of assimilation toward Stimulus 9, this middle stimulus will tend to be overestimated. Suppose it was called Stimulus 6. The feedback then informed the subject that the correct response was Stimulus 5. Now the subject has a problem. The error did not occur because the subject had been guessing wildly and had no confident assessment that the stimulus was Stimulus 6. Subjects are sometimes so confident of their response that they ask the experimenter if the feedback was correct. Also, they sometimes report that their response scale seems to shift about. They may report strings of trials in which they consistently overestimate or underestimate the stimuli. Indeed, some difficulty in keeping the response and stimulus scales aligned is expected. If assimilation is real, then Stimulus 5 really appeared like Stimulus 6 in memory, and the Stimulus 5 feedback can only mean that the response scale is miscalibrated (or that the feedback was wrong). A solution for the subject is to shift the response scale so as to realign it with the stimuli. In this example, the response scale should be shifted down one category unit. If it is, and if this is accomplished before the next stimulus, then the response on that trial will tend to be contrasted with Stimulus 9 two trials earlier. Obversely, if the first stimulus had been Stimulus 1 rather than Stimulus 9, the response scale would have shifted upward, and the response to the stimulus two trials later would tend toward large, producing contrast to the 1 on trial S_{N-2}.

Shifting the response scale after every trial might produce contrast to S_{N-2} on average, but a few trial examples quickly show that it would not produce contrast to every combination of prior stimuli. For example, it predicts assimilation two trials back whenever the three involved stimuli steadily increase or decrease in intensity. We do not find support for this in the data. Also, models incorporating only very recent trials (e.g., Holland & Lockhead, 1968; Wagner & Baird, 1981) predict a decrease in the magnitude of the sequential effect with increasing trials, but this is not the usual result. Usually, as in Figures 8 and 9, the magnitude of the contrast is uniform over many trials, or it peaks at trial $N - 3$. Perhaps the scale is not shifted on every trial. Rather, the shift appears to be in terms of some running average of previous events, like S_P or M_P. The suggestion is that perceived stimuli are continuously placed into the memory pool. If several recent stimuli had been loud, then assimilation also makes these appear loud. The subsequent memory pool thus tends to be large, and current stimuli are judged in terms of this pool. Feedback then informs the subject that the responses are too large; the subject thus shifts the response scale such that subsequent stimuli are responded to with smaller numbers. This produces contrast between the response and the events composing M_P. To model these speculations, consider that the subject adjusts the response scale by comparing the memory pool with the average of all memories, and adjusts the response in terms of this difference. This would require an added term to Equation 4 such that

$$R_N = S_N + a(M_{N-1} - S_N) + b(\bar{M} - M_P), \quad (8)$$

where M_P is the average of the recent memories (estimated here by trials $N - 2$ through $N - 6$), $\bar{M}$ is the average memory in the experiment, b is a positive constant, and the other terms are as described for Equation 4.

Equation 8 states that S_N is compared with the biased memory of the previous stimulus, M_{N-1}, and that a response is selected in terms of how the memory pool has been aligned with the response scale. If the memory pool at the instant of judgment is greater than the average pool, $M_P > \bar{M}$, then the response will be smaller than when the changing memory pool is less than average. This will produce contrast between the current stimulus value and the average of events prior to trial $N - 1$.

M_{N-1}, $\bar{M}$, and M_P in Equation 8 are hypothetical constructs and cannot be directly measured. However, R_{N-1} is an estimate of M_{N-1}; the average stimulus, $\bar{S}$, is an estimate of $\bar{M}$; and the average of the stimuli on trials $N - 2$ through $N - 6$ estimate M_P. If such estimates are reasonably good, then R_N in Equation 8 might be estimated by

$$R_N = S_N + a(R_{N-1} - S_N) + b(\bar{S} - S_P). \quad (9)$$

Table 1 shows the standardized regression weights obtained when Equation 9 was evaluated for each of the four subjects in a mag-

Table 1
Standardized Coefficients From a Model to Predict Responses in the Magnitude-Estimations-With-Feedback Experiments

Subject	k	a	b	R^2
1	1.077	.2058	−.0559	.8915
2	1.165	.4119	−.0835	.8389
3	1.079	.2237	−.0780	.8667
4	1.036	.1567	−.0756	.8842

Note. R^2 is the estimated proportion of variance in the responses that is accounted for by the model. The model evaluated is based on Equation 9 in the text and is as follows:

$$\log R_N = k \log S_N + a(\log R_{N-1} - \log S_N) + b(\log S_{N-2} + \cdots + \log S_{N-6})/5 + c.$$

nitude-estimations-with-feedback study using the same 30 stimuli as in the SRJ task here (King & Lockhead, 1981, Experiment 1). The current stimulus, *k*, contributes most to the response, as must occur whenever performance is good. The weight associated with the prior response, *a*, is consistently positive and significant, indicating assimilation, and the weight associated with the stimulus pool, *b*, is consistently negative and significant, indicating contrast (all $ps < .0001$).

Table 1 shows the results of an evaluation of Equation 9 when there was feedback. There, as in all feedback studies, *b* should be negative to describe the observed contrast to earlier events. When feedback is not given in the experiment, the equation is still appropriate, but *b* is expected to be small because veridical information concerning the pool is less available to the subject and because there is little evidence of contrast at trials $N - 2$ or $N - 3$. But *b* should still be real (and perhaps the pool analysis should go further but than six trials in no-feedback data) because there is usually contrast to events several trials back in no-feedback data (see Ward & Lockhead, 1970, Figures 2 and 4 for an AJ example; see Jesteadt et al., 1977, Figure 2, right panel, for an ME example, although they interpret those data differently). When there is no feedback, contrast may be due to response scales shifts that occur when subjects observe that they have not recently been using all available responses (cf. King, 1980).

This is surely not a complete description of the process involved. For example, there is no parameter for the effects of the time between trials, and the intertrial interval is important to identification performance and to sequential effects (Holland, 1968). However, this and other parameters that we found important could be introduced and the general approach may be promising. We know that a second-order linear equation, of which Equation 9 is one specific example, can describe assimilation and contrast (see Staddon et al., 1980, particularly their Equation 16). The particular model tested in that earlier article did not satisfactorily account for contrast several trials back in the data; Equation 8 (or 9) may accomplish this and may be a promising format for the study of memory effects in judgment tasks.

Discussion

Memory is a capacity for showing effects as the result of past treatment. Some sort of memory is thus involved whenever performance depends on past events, as is the case for the data discussed in this article. Absolute judgments, magnitude estimations, and successive-ratio judgments are not simply due to the subjects responding to individual stimuli as separate and independent things; they are not direct judgments. Just as in many other classification studies (for example: Broadbent & Ladefoged, 1967; Gravetter & Lockhead, 1973; Helson, 1948; Rosch, 1975), we can conclude that memory in these judgment tasks is important to performance.

This conclusion has not universally been held to be true. For example, it has been reported that enduring sequential effects are not real, that events remote to the stimulus are not important in AJ and ME data (Jesteadt et al., 1977). Such a conclusion is consistent with direct-scaling methods, which assume that "the observer is able to describe his observations at the quantitative level demanded by the instructions" (Engen, 1971, p. 64). Some caution is called for, however. Jesteadt et al.'s (1977) conclusion required the acceptance of a null hypothesis based on a problematic analysis (Staddon et al., 1980), and the statement that people can make direct observations is an assumption (Coombs, Dawes, & Tversky, 1970, pp. 18–19). In its simplest form, this assumption requires that "equal physical ratios produce equal subjective ratios" (Stevens, 1957, p. 153) and is the axiomatic basis of Stevens' law (see Equation 1).

This equality of ratios is often met in physics, as when weights are measured by means of a balance pan. It is also often supported by psychophysical data, as long as judgments are averaged over many trials and intensities. However, this axiom is not supported when individual responses are analyzed. Trial-to-trial analyses reveal a regular pattern that is inconsistent with an assumption of the equality of ratios. The example we focus on is that stimulus ratios of 1 are not usually reported as 1 (see Figure 3). Another difficulty is that small stimulus ratios tend to be overestimated, whereas large stimulus ratios are underestimated (see Figure 1). This observation is consistent with an earlier one that led Hollingworth (1910) to propose a central-tendency theory. Needham (1935) later suggested that this central-tendency result occurs because the remembered value of a standard shifts toward the average of the stimulus series, whereas Johnson (1952) and Stevens (1975, pp. 271–281) attributed this shift to the statistical regression of responses toward the series mean. However, Figures 5, 6, and 7 show that responses tend to shift *away* from the mean rather than toward it. This is the opposite of what would occur if central tendency is due to regression toward the mean. Apparently, the central tendency in Figure 1 is not due to regression but is the consequence of averaging many assimilations.

Trial-to-trial analysis of data other than successive-ratio judgments also show that the response shift is toward the previous trial, not toward the series midpoint. Absolute judgments tend to shift away from the series mean whenever the stimulus is intermediate to that mean and to S_{N-1} (Holland & Lockhead, 1968, Figure 4; Ward & Lockhead, 1971). This same result also occurs in magnitude-estimation data (King, 1980, Figures A5 and A6). It may be a general finding that central tendency is the result of averaging the many trial-to-trial memory effects modeled here.

Another example of the value of conducting trial-by-trial data analyses relates to a prior account to explain assimilation. Elmasian, Galambos, and Bernheim (1980, p. 606) concluded that assimilation is the result of successive stimuli merging within the sensory system, by means of some overlapping process, such that the stimuli become neurally confused with one another. The idea is that the neural structures involved when S_N occurs are also involved when S_{N+1} occurs, and that there is an averaging or summing of these representations such that successive stimuli are received as overly like one another. This is not consistent with the fact that successive identical stimuli are most often judged to be different rather than the same. A neural-confusion model must predict that physically identical objects are the ones most likely to be perceived as the same because these are the ones most likely to involve the same neural structures. That identical stimuli are not reported to be identical, whereas physically different events are reported as the same, suggests that assimilation is not simply the result of the neural merging of successive intensities.

These few observations show that averaged data can be misleading in the search for an account of how stimuli are processed. Responses to stimuli depend on the environment in which they occur. Thus, context must be included as a factor in the data analyses if the processes producing those data are to be understood. This suggests that judgment data should regularly be analyzed trial by trial so that context effects are not averaged across and, thus, so that contributions of memory to performance might be made visible.

References

Baker, K. E., & Dudek, F. J. Weight scales from ratio judgments and comparisons of existent weight scales. *Journal of Experimental Psychology,* 1955, *50,* 293–308.

Broadbent, D. E., & Ladefoged, P. Vowel judgments and adaptation level. *Proceedings of the Royal Society B,* 1967, *151,* 384–399.

Coombs, C. H., Dawes, R. M., & Tversky, A. *Mathematical psychology.* Englewood Cliffs, N.J.: Prentice-Hall, 1970.

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

Elmasian, R., Galambos, R., & Bernheim, A. Loudness enhancement and decrement in four paradigms. *Journal of the Acoustical Society of America,* 1980, *67,* 601–607.

Engen, T. Psychophysics: II. Scaling methods. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth and Schlosberg's experimental psychology.* New York: Holt, Rinehart & Winston, 1971.

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

Helson, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychological Review,* 1948, *55,* 297–313.

Holland, M. K. Channel capacity and sequential effects: The influence of the immediate stimulus history on recognition performance. (Doctoral dissertation, Duke University, 1968). *Dissertation Abstracts International,* 1969, *30,* 1923B. (University Microfilms No. 69-16,758)

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

Hollingworth, H. L. The central tendency of judgment. *Journal of Philosophy, Psychology, and Scientific Methods,* 1910, *7,* 461–468.

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

Johnson, D. M. The central tendency of judgment as a regression phenomenon. *American Psychologist,* 1952, *7,* 281.

King, M. C. Sequential effects in the variability of magnitude estimations of loudness. (Doctoral dissertation, Duke University, 1980). *Dissertation Abstracts International,* 1981, *41,* 3609B. (University Microfilms No. DEN 81-05068)

King, M. C., & Lockhead, G. R. On memory effects in magnitude estimation experiments. *Perception & Psychophysics,* 1981, *30,* 599–603.

Luce, R. D., & Green, D. M. The response ratio hypothesis for magnitude estimation. *Journal of Mathematical Psychology,* 1974, *11,* 1–14.

Metfessel. M. A. A proposal for quantitative reporting of comparative judgments. *Journal of Psychology,* 1947, *24,* 229–236.

Needham, J. G. Contrast effects in judgments of auditory intensities. *Journal of Experimental Psychology,* 1935, *18,* 214–236.

Newhall, S. M. The ratio method in the review of the Munsell colors. *American Journal of Psychology,* 1939, *52,* 394–405.

Richards, A. A non-metric scaling of loudness. *Journal of the Acoustical Society of America,* 1974, *56,* 582–588.

Richardson, L. F. Quantitative mental estimates of light and colour. *British Journal of Psychology,* 1929, *20,* 27–37.

Richardson, L. F., & Ross, S. J. Loudness and telephone current. *Jounal of General Psychology,* 1930, *3,* 288–306.

Rosch, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General,* 1975, *104,* 192–233.

Staddon, J. E. R., King, M. C., & Lockhead, G. R. On sequential effects in absolute judgment experiments. *Journal of Experimental Psychology: Human Perception and Performance,* 1980, *6,* 290–301.

Stevens, S. S. On the psychophysical law. *Psychological Review,* 1957, *64,* 153–181.

Stevens, S. S. *Psychophysics.* (G. Stevens, Ed.) New York: Wiley, 1975.

Wagner, M., & Baird, J. C. A quantitative analysis of sequential effects with numeric stimuli. *Perception & Psychophysics,* 1981, *29,* 359–364.

Ward, L. M., & Lockhead, G. R. Sequential effects and memory in category judgments. *Journal of Experimental Psychology,* 1970, *84,* 27–34.

Ward, L. M., & Lockhead, G. R. Response system processes in absolute judgment. *Perception & Psychophysics,* 1971, *9,* 73–78.

Received May 20, 1982
Revision received September 17, 1982 ■