Induction of combination rules
in two-dimensional function learning

KYUNGHEE KOH
University of Rochester, Rochester, New York

Previous studies have typically found that when people learn to combine two dimensions of a stimulus to select a response, they learn additive combination rules more easily than nonadditive (e.g., multiplicative) ones. The present experiments demonstrate that in some situations people can learn multiplicative rules more easily than other (e.g., additive) rules. Subjects learned to produce specified response durations when presented with stimulus lines varying in length and angle of orientation. When stimuli and correct responses were related by a multiplicative combination of power functions, learning was relatively easy (Experiment 1). In contrast, systematic response biases occurred during the early phases of learning an additive combination of linear functions (Experiment 2) and a more complex (nonadditive and nonmultiplicative) combination of linear functions (Experiment 3), suggesting that people have a tendency to induce a multiplicative combination of power functions. However, the initial biases decreased with practice. These results are explained in terms of a revised adaptive regression model of function learning originally proposed by Koh and Meyer (1991). Differences between the present results and previous results in the literature are discussed.

In many tasks, the appropriate response depends on several stimulus variables. For example, investment decisions should depend on risks and potential gains, and hitting a ball requires judging the velocity and distance of the ball. The question of how stimulus variables are combined has therefore received considerable attention in many different contexts, such as decision making (Tversky, 1967), measurement theory (Anderson, 1981; Luce & Tukey, 1964), intuitive physics (Anderson, 1983), and visual perception (Maloney & Landy, 1989).

The present article explores one important aspect of information combination—namely, learning of specific combination rules from experience—in the context of function learning. In a function learning paradigm, subjects learn quantitative functional relations among continuous stimulus and response variables (e.g., Bjorkman, 1965; Brehmer, 1974; Carroll, 1963; Koh & Meyer, 1991). Within the function learning paradigm, induction of combination rules can be investigated by examining how people learn multidimensional functions.

A multidimensional function, which relates multiple stimulus variables to a response variable, can be thought of as having two distinct parts. One specifies the component functions—that is, how the individual stimulus variables are related to the response variable when other stimulus variables are fixed (e.g., by a linear or power function). The other specifies the combination rule—that is, how the component functions are combined (e.g., additively or multiplicatively). Previous studies of function learning have typically concluded that (1) linear component functions are learned faster than nonlinear (e.g., U-shaped) ones (Brehmer, 1974; Carroll, 1963; Deane, Hammond, & D. A. Summers, 1972; Hammond & D. A. Summers, 1965), and that (2) additive combination rules are learned more easily than nonadditive (e.g., multiplicative) ones (Brehmer, 1969; S. A. Summers, R. A. Summers, & Karkau, 1969). (For reviews, see Brehmer, 1980, and Klayman, 1988.)

These conclusions, however, are based mainly on studies where numbers or line lengths were used as stimulus and response variables, and they may not necessarily generalize to learning of relations among other variables. Indeed, with respect to the learnability of component functions, the recent work of Koh and Meyer (1991) has shown that under some circumstances power functions can be learned more easily than linear functions. Koh and Meyer had subjects learn to select particular response durations when presented with particular stimulus lengths. Stimuli and correct responses were related by either a power function, a logarithmic function, or a linear function with a positive intercept. The power function was learned quickly and accurately. In contrast, systematic response biases occurred during the early phases of learning the logarithmic and linear functions, although the biases gradually decreased with practice. The pattern of initial response biases suggested that subjects made an a priori assumption that the function to be learned was a power function.

On the one hand, these results are surprising given that several other studies of function learning have concluded that linear functions are more easily learned than nonlinear ones (Brehmer, 1974; Carroll, 1963; Deane et al., 1972; Hammond & D. A. Summers, 1965). On the other hand, these results mesh well with the findings of numerous psychophysical scaling studies that have found power psychophysical functions for many sensory dimensions (S. S. Stevens, 1956, 1957, 1961). Many of these psychophysical studies have used cross-modality matching tasks (J. C. Stevens, Mack, & S. S. Stevens, 1960; J. C. Stevens & Marks, 1965; S. S. Stevens, 1959, 1965; S. S. Stevens & Guirao, 1963), which are similar to typical function learning tasks except that subjects choose their own response to each stimulus rather than having to learn a predefined response. The results of cross-modality matching tasks therefore could be interpreted as showing the special status of power functions in relating sensory dimensions.

One way to reconcile the apparent discrepancy between the results showing the primacy of linear functions and those showing the primacy of power functions is to note that power functions are closely related to linear functions. A logarithmic transformation of stimulus and response variables turns power functions into linear functions. Thus, the primacy of both linear functions and power functions may be interpreted as indicating that induction of linear functions is preferred over induction of nonlinear functions, with the exception that an optional logarithmic transformation of stimulus and response variables sometimes precedes function induction. Whether or not a logarithmic transformation of variables takes place in a particular situation may depend on the nature of the stimulus and response variables involved. (This will be discussed in more detail in the General Discussion section.)

These considerations have important implications for the learnability of combination rules. Suppose that induction of additive combination rules is preferred over induction of nonadditive rules, with the exception that an optional logarithmic transformation of stimulus and response variables sometimes precedes the induction process. Then it is possible that multiplicative rules (i.e., rules that are additive after logarithmic rescaling) may be learned more easily than other (e.g., additive) rules. The present experiments provide a demonstration that in some situations people can learn multiplicative rules more easily than additive rules.

In the present three experiments, the subjects learned to produce specified response durations when presented with stimulus lines varying in length and angle of orientation.[1] Two of the three dimensions used here (length and duration) were the same as those used in Koh and Meyer (1991). The correct response durations were related to the values along the length and angle dimensions by a multiplicative combination of power functions, an additive combination of linear functions, or a more complex (nonadditive and nonmultiplicative) combination of

linear functions. The experiments were designed to test the hypothesis that in learning to relate line length and angle to response duration, subjects have a natural bias to assume that the function to be learned is a multiplicative combination of power functions and that they can overcome the bias so as to learn other types of function. According to the hypothesis, subjects should be able to learn a multiplicative combination of power functions relatively quickly and accurately, whereas they should show systematic response biases toward a multiplicative combination of power functions in learning other types of function.

EXPERIMENT 1

In Experiment 1, the subjects learned a two-dimensional function in which power component functions were combined multiplicatively. The aim was to test the hypothesis that people can learn the required multiplicative combination of power functions quickly and accurately.

Method

Subjects. Five University of Rochester students participated as paid subjects. Each subject received $5.00 per 1-h session, plus a bonus of about $0.50 on the basis of performance. The subjects were tested in three 1-h sessions on separate days over a period of a week or less.

Apparatus. An Apple Macintosh IIci microcomputer controlled the presentation of stimuli and collection of responses. The subjects sat in a normally lit room and viewed the CRT screen of an AppleColor high-resolution RGB monitor from a distance of approximately 70 cm. Responses were made through a standard computer keyboard. A software timing routine with a resolution of 12 msec measured the stimulus and response durations.[2]

Design. The stimuli were lines varying in length and angle of orientation.[3] Five levels of length combined with five levels of angle produced a set of 25 stimuli. The levels were chosen so that the stimuli could be discriminated easily. The function that related the stimuli to the correct response durations was $D = 71.06 \, L^{.32} A^{.32}$, where D, L, and A denote the duration (in milliseconds), length (in millimeters), and angle (in degrees from horizontal), respectively. Table 1 lists the stimulus lengths, stimulus angles, and response durations used in Experiment 1.

Each session began with a block of 10 warm-up trials for which the data were not analyzed. The warm-up trials were followed by 12 50-trial blocks. During each block, each of the 25 stimuli was presented twice. The order of stimulus presentation within a block was randomized. Feedback about the correct responses was provided on 32 randomly selected trials during each block.[4]

Procedure. At the beginning of each trial, a line was presented at the center of the display screen. The subjects were instructed to produce a response duration that they thought would correspond to the length and angle of the stimulus line. They were told that there was a correct response duration for each stimulus, but they were not told explicitly about the nature of the underlying stimulus-response relation. The response duration was defined as the amount of time between two taps of the zero ("0") key on the keyboard.

A warning message appeared on the display screen if a procedurally inappropriate response occurred. The inappropriate responses included responding before stimulus onset, not responding within 5 sec after stimulus onset, and not tapping the zero key. The warning messages were, respectively, "Do not respond before stimulus presentation," "You did not respond before dead-

Table 1
The Stimulus-Response (S-R) Pairs in Experiment 1

S-R Pair	Stimulus Length (mm)	Stimulus Angle (degree)	Response Duration (msec)
1	13.9	9	333
2	13.9	25	462
3	13.9	41	541
4	13.9	57	600
5	13.9	73	650
6	38.2	9	461
7	38.2	25	639
8	38.2	41	748
9	38.2	57	831
10	38.2	73	899
11	62.4	9	540
12	62.4	25	748
13	62.4	41	876
14	62.4	57	973
15	62.4	73	1053
16	86.7	9	600
17	86.7	25	831
18	86.7	41	973
19	86.7	57	1081
20	86.7	73	1170
21	111.0	9	650
22	111.0	25	900
23	111.0	41	1054
24	111.0	57	1171
25	111.0	73	1267

line," and "You did not tap the '0' key." Each warning message lasted 750 msec, and after an intertrial interval (ITI) of 830 msec, the next trial began. Trials involving inappropriate responses were not repeated.

On feedback trials, if the subject responded appropriately, various types of feedback were provided. First, two brief beeps whose onsets were separated by the correct response duration were presented. The first beep started 330 msec after the subject's second keytap. After the second beep, some further information about the subject's performance was presented on the display screen. Depending on whether the response duration was longer than, shorter than, or equal to the correct duration, the message "Long," "Short," or "Perfect" appeared on the screen. Below the message, a point score for the response was shown. The point score ranged from 0 to 10, indicating how close the subject's response duration had come to the correct duration.[5] The message "Perfect" was presented when the score for the response was 10. The feedback message lasted 1 sec, and the next trial began after an 830-msec ITI.

On trials without feedback, if the subject made a procedurally appropriate response, an ITI of 830 msec followed the second keytap, and the next trial began. The subjects were told that on some randomly selected trials no feedback would be given, and that whether or not they received feedback on a trial did not depend on their performance on the trial.

During a brief rest period between blocks, the subjects received information about the number of points that they had earned for the previous block and the cumulative number of points for the session. The next block of trials then began. At the end of each session, the subjects received a bonus payment of a penny for 10 points above 3,000 points.

Data analyses. Each subject produced a total of 1,800 response durations (3 sessions × 12 blocks × 50 trials). A small number of procedurally inappropriate responses were excluded from data analyses. All analyses were based on logarithmically transformed stimulus and response values.

First, a mean and standard deviation of log response durations was computed for each combination of subject, stimulus-response

pair, and half-session. Observations were grouped into half-sessions rather than into sessions, because a preliminary analysis suggested that a substantial amount of learning took place within the first session. Each mean and standard deviation was based on about 12 observations, which were enough to produce fairly stable estimates. The mean response (or the difference between the mean and correct responses, i.e., constant error) provides a measure of systematic bias, whereas the standard deviation (i.e., variable error) provides a measure of response consistency. The two performance measures reflect different sources of error and thus are analyzed separately.

In order to examine the pattern of subjects' responses, the means of log response durations [$M(\log D)$] were used as the variable to be predicted in multiple regression analyses. The prediction equation was derived from the following log-quadratic equation:

$$M(\log D) = a_1 + a_2 \log L + a_3 \log A + a_4(\log L)(\log A)$$
$$+ a_5(\log L)^2 + a_6(\log A)^2, \qquad (1)$$

where D, L, and A denote the response duration, stimulus length and stimulus angle, and a_1 through a_6 represent the regression coefficients. Because the predictor variables in Equation 1 are correlated, a slightly modified form of Equation 1 was used. The set of correlated predictor variables was transformed into another set of normalized orthogonal variables. This was done in several steps. The variables $\log L$ and $\log A$ were first standardized so that the new variables (Z_L and Z_A) had means of 0 and standard deviations of 1. Standardization was done to make the magnitudes of the different regression coefficients comparable. Next, Z_L and Z_A were used to predict each of the remaining three variables, ($\log L$)($\log A$), ($\log L$)^2, and ($\log A$)^2, and then the standardized residuals, Z_{LA}, Z_{LL}, and Z_{AA}, were computed. These transformations removed linear trends from the original variables, ($\log L$)$(\log A$), ($\log L$)^2, and ($\log A$)^2. As a result, the coefficients for the new variables (Z_{LA}, Z_{LL}, and Z_{AA}) reflected pure second-order trends. With these new orthogonal predictor variables, the regression equation became

$$M(\log D) = c_1 + c_2 Z_L + c_3 Z_A + c_4 Z_{LA}$$
$$+ c_5 Z_{LL} + c_6 Z_{AA}. \qquad (2)$$

This equation was chosen because it can approximate a variety of two-dimensional functions (including multiplicative combinations of power functions and additive combinations of linear functions) with a relatively small number of parameters (c_1 through c_6). Furthermore, given the logarithmic transformations of the stimulus and response variables, the equation reduces to a multiplicative combination of power functions when c_4, c_5, and c_6 equal zero. Thus, if responses are based on a multiplicative combination of power functions, the estimates of the coefficients c_4, c_5, and c_6 should be essentially zero. A nonzero value of c_4 would indicate an interaction between the log stimulus variables (i.e., deviation from a multiplicative combination rule), and nonzero values of c_5 and c_6 would indicate log-quadratic trends in the component functions (i.e., deviations from power functions).[6]

Regression coefficients were computed separately for each combination of subject and half-session. Each of the six sets of estimated coefficients (one for each term in Equation 2) was then submitted to an analysis of variance (ANOVA), with half-sessions as a within-subject fixed factor and subjects as a random factor.

Finally, variable errors (i.e., standard deviations of log response durations) were submitted to an ANOVA, with half-sessions as a within-subject fixed factor and subjects as a random factor.

Results

Overall stimulus-response relation. Logarithmically transformed response durations, averaged over subjects, appear in Table 2 for each combination of stimulus-response pair and half-session, along with the correct log

Table 2
Mean Log Response Durations for Each Stimulus-Response (S-R) Pair Averaged Over Subjects as a Function of Half-Session and the Correct Log Response Duration for Each Pair in Experiment 1

S-R Pair	Correct Log Response	Half-Session					
		1	2	3	4	5	6
1	5.808	5.897	5.837	5.879	5.805	5.920	5.900
2	6.136	6.024	6.094	6.101	6.119	6.122	6.150
3	6.293	6.169	6.228	6.281	6.316	6.278	6.290
4	6.397	6.141	6.290	6.294	6.316	6.368	6.370
5	6.477	6.302	6.366	6.433	6.491	6.468	6.382
6	6.133	6.194	6.173	6.149	6.164	6.225	6.196
7	6.460	6.421	6.484	6.496	6.459	6.508	6.544
8	6.617	6.562	6.583	6.612	6.669	6.629	6.648
9	6.723	6.610	6.635	6.687	6.704	6.711	6.737
10	6.801	6.695	6.763	6.791	6.799	6.812	6.785
11	6.292	6.390	6.329	6.281	6.295	6.315	6.327
12	6.617	6.627	6.563	6.650	6.699	6.659	6.668
13	6.775	6.661	6.665	6.758	6.757	6.785	6.758
14	6.880	6.809	6.775	6.832	6.751	6.849	6.795
15	6.959	6.895	6.931	6.947	6.909	6.927	6.862
16	6.397	6.456	6.469	6.324	6.444	6.405	6.425
17	6.723	6.710	6.710	6.783	6.765	6.752	6.739
18	6.880	6.830	6.795	6.825	6.856	6.873	6.822
19	6.986	6.932	6.839	6.852	6.854	6.941	6.888
20	7.065	7.005	6.989	7.043	6.989	7.009	6.960
21	6.477	6.618	6.595	6.518	6.595	6.530	6.587
22	6.802	6.837	6.786	6.851	6.846	6.872	6.881
23	6.960	6.888	6.910	6.950	6.974	6.981	6.947
24	7.066	7.015	7.013	7.027	6.977	7.020	6.995
25	7.144	7.081	7.082	7.103	7.077	7.129	7.083

responses. In Figure 1, the results from the first half of Session 1 (top panels) and the second half of Session 3 (bottom panels) are shown. The left panels present the log response durations as a function of log stimulus lengths with angle as a parameter, and the right panels present the same information as a function of log stimulus angles with length as a parameter. The dotted parallel lines represent the required multiplicative combination of power functions, and the circles on the solid curves (connected line segments) represent the observed responses.

Performance during the first half of Session 1 did not differ much from performance during the second half of Session 3, suggesting that the subjects attained a nearly asymptotic level of performance during the first half-session. The subjects' responses came close to the required ones, except that the observed response magnitudes for the short and steep stimulus lines tended to be smaller than required. Overall, the response curves (solid curves) were almost linear and parallel to each other, consistent with the hypothesis that the subjects' responses were based on a multiplicative combination of power functions. Detailed multiple regression analyses provide support for these general impressions.

Multiple regression coefficients. Multiple regression analyses were performed using Equation 2 as the prediction equation. There were a total of 30 such regression analyses, one for each combination of 5 subjects and 6

half-sessions. The fits were uniformly good; the percentage of variance accounted for was 96.8%, averaged over subjects. For individual subjects, the percentages ranged from 95.9% to 98.2%.

Figure 2 presents the regression coefficients, averaged over subjects, as a function of half-sessions. Each of the six panels shows the obtained coefficients (circles) for each term in Equation 2. Each solid line represents the ideal value of the coefficient that would result from perfect performance.

Several features of the results are noteworthy. First, the intercepts (c_1) were initially smaller than the ideal, but approached the ideal with practice. Their increase over half-sessions was statistically reliable $[F(5,20) = 3.84, MS_e = .00052, p < .02]$. Second, the log-linear coefficients for length (c_2) were overall not significantly different from the ideal $[F(1,4) = 1.71, MS_e = .00088, p > .2]$. There was a slight improvement with practice, but it was not statistically reliable $[F(5,20) = 1.47, MS_e = .00095, p > .2]$. Third, the log-linear coefficients for angle (c_3) were generally smaller than the ideal $[F(1,4) = 8.98, MS_e = .00689, p < .05]$.

The remaining three coefficients $(c_4, c_5,$ and $c_6)$ were not reliably different from zero $[F(1,4) < 1, MS_e = .0013; F(1,4) < 1, MS_e = .00056; F(1,4) < 1, MS_e = .0017]$, and they did not change significantly over half-sessions $[F(5,20) = 2.16, MS_e = .0036, p > .1; F(5,20) < 1, MS_e = .00015; F(5,20) = 2.39, MS_e = .0003, p > .05]$. The percentage of variance accounted for by these three coefficients was, on average, only 1.3%, in contrast to 95.5% accounted for by the first three coefficients. Recall that c_4 represents how much the combination rule deviates from a multiplicative (log-additive) rule, and that c_5 and c_6 represent how much the component functions deviate from power functions. Clearly, the subjects' responses were well fit by a multiplicative combination of power functions.

Variable errors. Variable errors averaged over subjects appear in Table 3 as a function of stimulus-response pair and half-session. There was a significant improvement in variable error with practice $[F(5,20) = 21.34, MS_e = .009, p < .0001]$. The mean variable errors were .261, .205, .177, .174, .151, and .167, for the six half-sessions. The variable errors also differed reliably across individual stimulus-response pairs $[F(24,96) = 2.89, MS_e = .003, p < .0001]$. In general, stimulus-response pairs with long response durations yielded smaller variable errors than did those with short response durations. The same pattern was observed in Koh and Meyer's (1991) unidimensional function learning studies, and it may reflect a general relationship between the mean and standard deviation of logarithmically transformed response durations.

Discussion

Overall, the results show that people can learn quite well a multiplicative combination of power functions re-
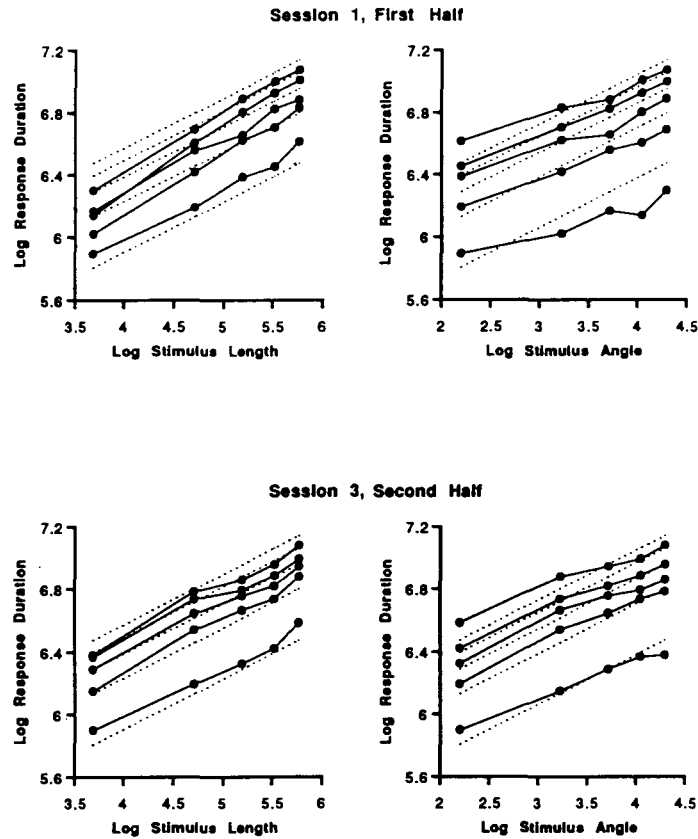
Figure 1. Mean log response durations (natural logarithms of milliseconds) averaged over subjects as a function of log stimulus lengths (natural logarithms of millimeters) and log stimulus angles (natural logarithms of degrees from horizontal) for the first half of Session 1 and the second half of Session 3 of Experiment 1. (The dotted lines represent the required function, and the circles on the solid curves represent the observed mean log response durations.)

lating stimulus length and angle to response duration. The subjects' responses conformed to a multiplicative combination of power functions even during the early stages of learning. Performance improved substantially in terms of variable error. There was also a modest improvement in mean responses with practice, but the improvement was attributable to better estimation of parameters for a multiplicative combination of power functions rather than to a change in the type of function being induced.

The present results replicate Koh and Meyer's (1991) finding that people readily learn a power-function relation between length and duration. The results also demonstrate that people can easily induce a power-function relation between angle and duration. These results, however, are surprising in light of the widely accepted view that additive combinations of linear functions are most easily learned (Brehmer, 1969; S. A. Summers et al., 1969). The results suggest that the type of combination rule and multidimensional function that is most easily learned may depend on the particular function learning task involved.

It is worth noting that, somewhat unexpectedly, the effects of length and angle were learned at different rates,

even though the subjects were explicitly told to pay attention to both angle and length. Whereas the obtained coefficients for length came close to the ideal within the first session, the coefficients for angle remained smaller than the ideal after three sessions of practice. It may be that the lengths used in the present experiment were more discriminable than the angles. Another possibility is that the two dimensions were equally discriminable, but in learning a two-dimensional function, people have a tendency to pay more attention to one stimulus variable than the other, instead of distributing attention equally across the two stimulus variables. More generally, people may have a tendency to focus on a few stimulus variables when there are many potentially relevant variables (cf. Busemeyer, Myung, & McDaniel, in press; Wallsten, 1976). Interestingly, the rather puzzling finding that the observed coefficients for angle were too small even after extensive training may have a simple statistical explanation. The relevant statistical fact is that errors of measurement in independent variables result in regression coefficients that are biased toward zero (Hays, 1988, pp. 583-584). Thus, the asymptotic coefficients for angle may have been too
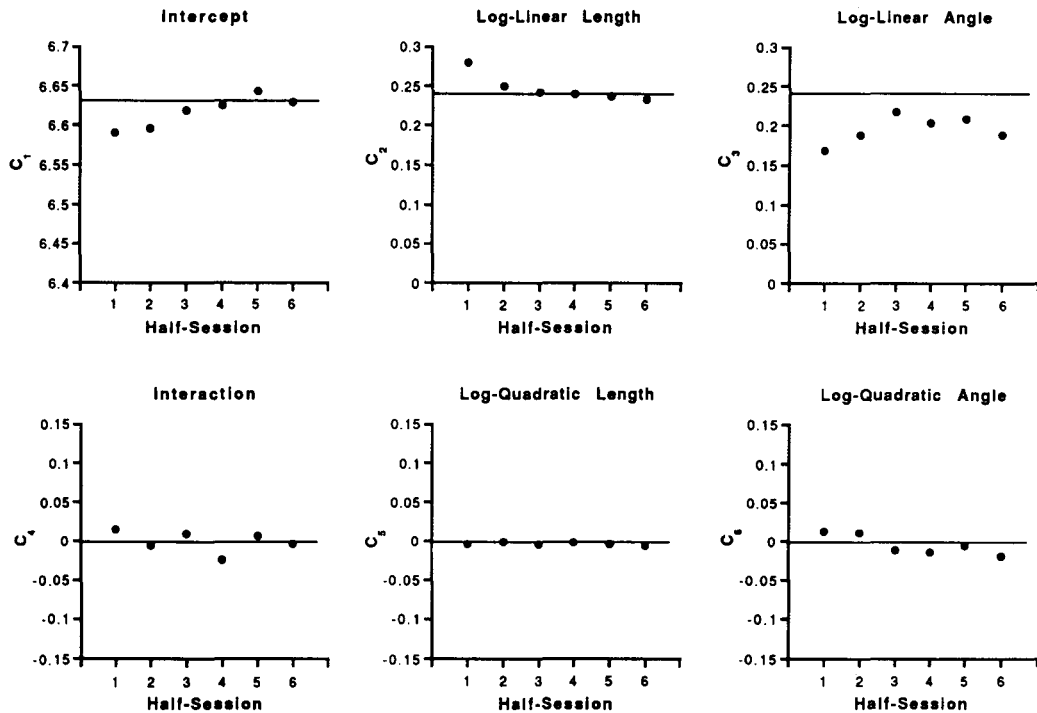
Figure 2. Multiple regression coefficients averaged over subjects as a function of half-sessions in Experiment 1. (The solid lines represent the ideal values of the coefficients.)

small in the present experiment because the internal representations of the angles were highly variable, either due to lack of attention or due to perceptual noise. This issue is discussed further in the General Discussion section.

EXPERIMENT 2

The results of Experiment 1 showed that a multiplicative combination of power functions, which is additive and linear in logarithmic coordinates, could be learned quite well. Experiment 2 was similar to Experiment 1, except that the required function was an additive combination of linear functions, which is neither additive nor linear in logarithmic coordinates. Thus, if subjects assume that the required function is a multiplicative combination of power functions, the results would initially show systematic response biases. Also, if subjects are able to overcome their initial biases, as in unidimensional function learning (Koh & Meyer, 1991), then they would be able to learn the present additive combination of linear functions with sufficient practice.

Method

The experimental method was the same as that of Experiment 1, except that the required rule was defined by the equation $D = 201 + 4.8L + 7.29A$, where D, L, and A denote the duration (in milliseconds), length (in millimeters), and angle of orientation (in degrees). The stimulus lengths, angles, and response durations are

Table 3
Standard Deviations of Log Response Durations for Each Stimulus-Response (S-R) Pair Averaged Over Subjects as a Function of Half-Session in Experiment 1

S-R Pair	Half-Session					
	1	2	3	4	5	6
1	0.313	0.196	0.191	0.188	0.189	0.174
2	0.221	0.201	0.195	0.167	0.178	0.177
3	0.273	0.208	0.164	0.164	0.156	0.154
4	0.192	0.164	0.168	0.164	0.149	0.164
5	0.205	0.188	0.153	0.183	0.145	0.217
6	0.374	0.224	0.160	0.179	0.159	0.154
7	0.346	0.280	0.169	0.196	0.143	0.185
8	0.337	0.216	0.206	0.165	0.169	0.141
9	0.342	0.203	0.202	0.231	0.185	0.195
10	0.346	0.193	0.165	0.143	0.161	0.151
11	0.257	0.242	0.180	0.215	0.190	0.168
12	0.211	0.168	0.170	0.192	0.161	0.179
13	0.195	0.174	0.184	0.168	0.179	0.176
14	0.230	0.156	0.161	0.165	0.132	0.166
15	0.191	0.229	0.151	0.155	0.138	0.144
16	0.309	0.212	0.212	0.147	0.150	0.150
17	0.231	0.188	0.177	0.237	0.172	0.198
18	0.262	0.230	0.168	0.190	0.137	0.159
19	0.236	0.185	0.171	0.169	0.146	0.149
20	0.192	0.206	0.205	0.145	0.183	0.149
21	0.302	0.213	0.196	0.192	0.154	0.172
22	0.223	0.195	0.177	0.161	0.163	0.153
23	0.256	0.233	0.168	0.145	0.173	0.143
24	0.267	0.208	0.179	0.151	0.140	0.170
25	0.214	0.218	0.156	0.151	0.125	0.158

Session 1, First Half
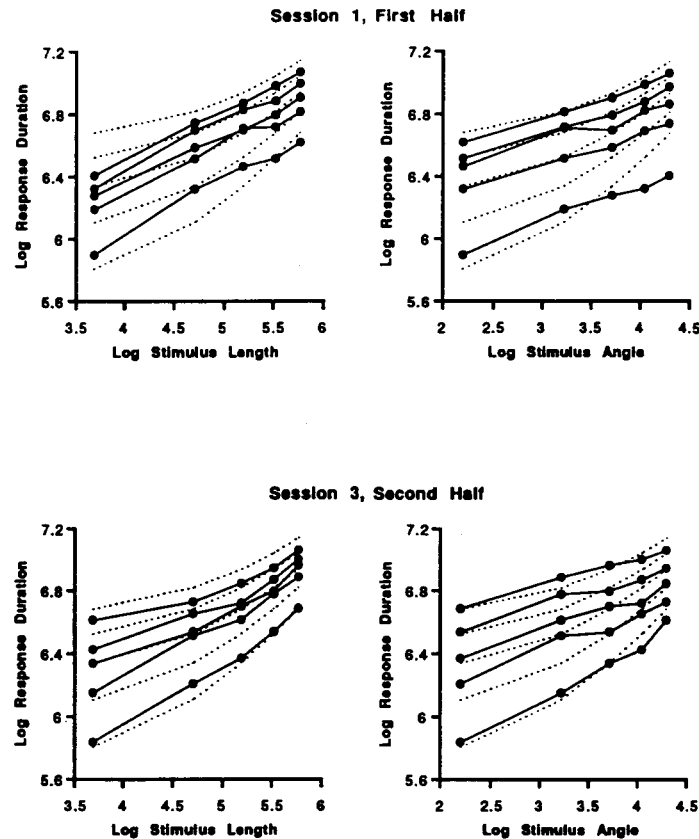

Session 3, Second Half

Figure 3. Mean log response durations (natural logarithms of milliseconds) averaged over subjects as a function of log stimulus lengths (natural logarithms of millimeters) and log stimulus angles (natural logarithms of degrees from horizontal) for the first half of Session 1 and the second half of Session 3 of Experiment 2. (The dotted curves represent the required function, and the circles on the solid curves represent the observed mean log response durations.)

listed in Table 4. The stimuli were the same as in Experiment 1, and the responses spanned the same range as those in Experiment 1.

Five University of Rochester students participated as paid subjects. Each subject received about $5.50 per 1-h session. None had been in Experiment 1.

Results

Overall stimulus–response relation. Log response durations, averaged over subjects, are listed in Table 5 as a function of stimulus-response pairs and half-sessions. The results from the first half of Session 1 and the last half of Session 3 also appear in Figure 3 as a function of stimulus lengths (left panels) and as a function of stimulus angles (right panels). The circles on the solid curves represent the observed response durations, and the dotted curves represent the required additive combination of linear functions.

During the first half of Session 1, there occurred interesting response biases. Note that the required additive combination of linear functions forms nonparallel curves (dotted curves) when plotted in logarithmic coordinates. In contrast, the observed responses (solid curves) formed approximately parallel lines. It appears that the subjects were initially inducing a multiplicative combination of power functions when an additive combination of linear functions was supposed to be learned. The response biases decreased substantially with practice, and by the second half of Session 3, the observed response durations came reasonably close to the correct ones. These observations are confirmed by detailed quantitative analyses.

Multiple regression coefficients. Multiple regression analyses and ANOVAs were performed as in Experiment 1. The regression fits were good; the percentage of variance accounted for, averaged over subjects, was 95.7%. For individual subjects, the percentages ranged from 92.6% to 98.4%. The regression coefficients, averaged over subjects, appear as a function of half-sessions in Figure 4. Solid lines represent the ideal values of the coefficients. These ideal values were computed by fitting Equation 2 to the correct log response durations. The regression fit accounts for 99.9% of the variance of the correct log responses, indicating that Equation 2 provides a good approximation to the required additive combination of linear functions. The dotted lines in the bottom panels represent zero.

The obtained intercepts (c_1) and log-linear coefficients for length (c_2) were on average not significantly differ-

Table 4
Stimulus-Response (S-R) Pairs in Experiment 2

S-R Pair	Stimulus Length (mm)	Stimulus Angle (degree)	Response Duration (msec)
1	13.9	9	333
2	13.9	25	450
3	13.9	41	567
4	13.9	57	683
5	13.9	73	800
6	38.2	9	450
7	38.2	25	567
8	38.2	41	683
9	38.2	57	800
10	38.2	73	917
11	62.4	9	567
12	62.4	25	683
13	62.4	41	800
14	62.4	57	917
15	62.4	73	1033
16	86.7	9	683
17	86.7	25	800
18	86.7	41	917
19	86.7	57	1033
20	86.7	73	1150
21	111.0	9	800
22	111.0	25	917
23	111.0	41	1033
24	111.0	57	1150
25	111.0	73	1267

ent from the ideal $[F(1,4) < 1, MS_e = .00064; F(1,4) < 1, MS_e = .00064]$. Unlike in Experiment 1 (where these coefficients were initially different from the ideal), in the present experiment, their initial values were quite close to the ideal. This difference may simply be due to individual differences, since there were a relatively small number of subjects in each experiment. As in Experiment 1, the log-linear coefficients for angle (c_3) were generally smaller than the ideal $[F(1,4) = 7.69, MS_e = .00797, p < .06]$, and although they improved somewhat with practice, this improvement was not reliable $[F(5,20) < 1, MS_e = .0012]$.

Unlike in Experiment 1, the ideal values for the latter three coefficients $(c_4, c_5,$ and $c_6)$ are different from zero. Yet, during the first half-session, the obtained coefficients were not reliably different from zero $[F(1,4) < 1, MS_e = .00098; F(1,4) < 1, MS_e = .00042; F(1,4) < 1, MS_e = .00269]$, and the deviations from the ideal values were more substantial $[F(1,4) = 10.59, MS_e = .00098; p < .05; F(1,4) = 6.19, MS_e = .00042, p < .1; F(1,4) = 3.128, MS_e = .00269, p < .2]$. The coefficients for the interaction between log length and log angle (c_4) approached the ideal with practice $[F(5,20) = 3.98, MS_e = .00031, p < .02]$, indicating that the subjects gradually learned the correct combination rule. Similarly, the log-quadratic coefficients for length (c_5) approached the ideal over half-sessions $[F(5,20) = 5.24, MS_e = .00026, p <$
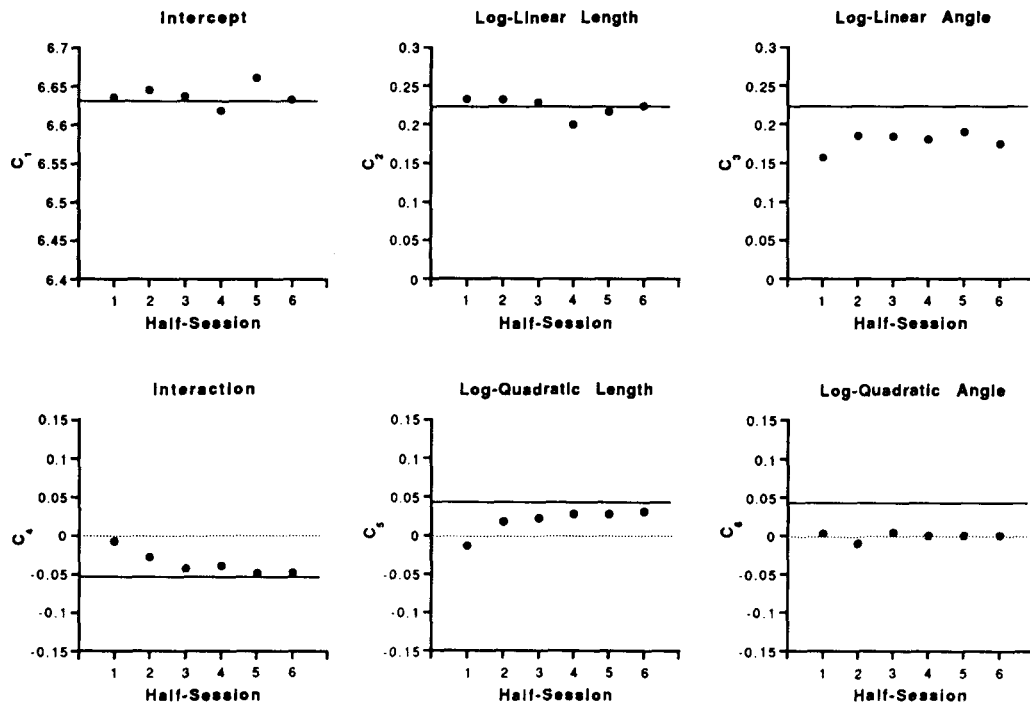
Figure 4. Multiple regression coefficients averaged over subjects as a function of half-sessions in Experiment 2. (The solid lines represent the ideal values of the coefficients, and the dotted lines represent zero.)

.005], which shows that subjects could learn the correct component function with respect to stimulus length. The log-quadratic coefficients for angle (c_6), however, did not change over half-sessions [$F(5,20) < 1$, $MS_e = .00019$].

Variable errors. Table 6 presents the average variable error for each half-session and stimulus–response pair. As in Experiment 1, the variable errors decreased somewhat over half-sessions [$F(5,20) = 2.60$, $MS_e = .062$, $p < .06$]. The average variable errors for the six half-sessions were .252, .187, .184, .189, .159, and .150. In addition, the variable errors differed somewhat depending on the stimulus–response pair involved [$F(24,96) = 1.64$, $MS_e = .004$, $p < .05$]. On the whole, the magnitudes of the variable errors in Experiment 2 were comparable to those in Experiment 1. The average variable errors for Experiments 1 and 2 were .189 and .187, respectively. It appears that even though the pattern of response biases (or constant errors) was affected by the form of the required function, response variability (or variable error) was largely unaffected by it.

Discussion

Experiment 2 produced several important results. First, even though the subjects were trained on an additive combination of linear functions, their initial responses conformed to a multiplicative combination of power functions. This result strongly suggests that in learning a function relating line length and angle to duration, peo-

Table 6
Standard Deviations of Log Response Durations for Each Stimulus–Response (S–R) Pair Averaged Over Subjects as a Function of Half-Session in Experiment 2

S–R Pair	Half-Session					
	1	2	3	4	5	6
1	0.268	0.213	0.181	0.182	0.226	0.151
2	0.284	0.212	0.232	0.193	0.173	0.154
3	0.315	0.204	0.220	0.218	0.131	0.148
4	0.348	0.226	0.231	0.155	0.175	0.148
5	0.353	0.218	0.190	0.204	0.153	0.139
6	0.269	0.197	0.164	0.189	0.161	0.141
7	0.227	0.197	0.186	0.225	0.180	0.154
8	0.246	0.132	0.126	0.158	0.159	0.156
9	0.253	0.175	0.210	0.180	0.135	0.155
10	0.224	0.196	0.159	0.215	0.135	0.132
11	0.237	0.214	0.189	0.215	0.133	0.173
12	0.244	0.192	0.171	0.195	0.152	0.140
13	0.193	0.260	0.180	0.188	0.158	0.133
14	0.208	0.140	0.147	0.211	0.150	0.167
15	0.195	0.138	0.178	0.189	0.165	0.142
16	0.330	0.214	0.201	0.161	0.172	0.143
17	0.286	0.183	0.196	0.175	0.173	0.153
18	0.190	0.161	0.148	0.136	0.156	0.157
19	0.187	0.162	0.171	0.185	0.178	0.133
20	0.231	0.151	0.180	0.223	0.137	0.138
21	0.256	0.219	0.205	0.197	0.165	0.163
22	0.244	0.177	0.174	0.197	0.145	0.154
23	0.217	0.166	0.201	0.189	0.123	0.151
24	0.262	0.176	0.195	0.154	0.153	0.182
25	0.242	0.163	0.172	0.197	0.179	0.152

Table 5
Mean Log Response Durations for Each Stimulus–Response (S–R) Pair Averaged Over Subjects as a Function of Half-Session and the Correct Log Response Duration for Each Pair in Experiment 2

S–R Pair	Correct Log Response	Half-Session					
		1	2	3	4	5	6
1	5.808	5.899	5.839	5.820	5.892	5.838	5.841
2	6.109	6.193	6.179	6.136	6.184	6.201	6.153
3	6.340	6.281	6.358	6.332	6.339	6.393	6.339
4	6.527	6.326	6.413	6.480	6.446	6.495	6.427
5	6.685	6.410	6.526	6.540	6.624	6.634	6.614
6	6.109	6.323	6.179	6.212	6.162	6.231	6.210
7	6.340	6.519	6.536	6.508	6.503	6.502	6.516
8	6.527	6.590	6.598	6.604	6.605	6.640	6.540
9	6.685	6.699	6.649	6.663	6.696	6.690	6.655
10	6.821	6.748	6.711	6.757	6.688	6.754	6.730
11	6.340	6.467	6.421	6.397	6.346	6.379	6.372
12	6.527	6.714	6.711	6.628	6.579	6.609	6.616
13	6.685	6.702	6.741	6.695	6.655	6.731	6.702
14	6.821	6.831	6.787	6.759	6.720	6.825	6.722
15	6.940	6.874	6.872	6.891	6.845	6.888	6.849
16	6.527	6.518	6.495	6.523	6.501	6.532	6.540
17	6.685	6.722	6.766	6.720	6.731	6.772	6.779
18	6.821	6.798	6.837	6.811	6.799	6.825	6.799
19	6.940	6.888	6.895	6.820	6.844	6.853	6.872
20	7.048	6.984	7.019	6.973	6.895	7.009	6.947
21	6.685	6.622	6.645	6.666	6.632	6.691	6.687
22	6.821	6.818	6.855	6.916	6.818	6.907	6.889
23	6.940	6.910	6.941	6.940	6.917	6.987	6.967
24	7.048	6.997	7.049	7.045	6.975	7.051	7.004
25	7.144	7.070	7.121	7.113	7.052	7.079	7.064

ple make an a priori assumption that the required combination rule is multiplicative and that the required component functions are power functions.

Second, performance improved with practice, suggesting that the underlying inductive mechanism is flexible enough to learn different types of two-dimensional function. In particular, the patterns of observed regression coefficients suggest that the combination rule that the subjects used shifted from a multiplicative rule to an additive rule, and that the subjects could learn the required nonpower function relating stimulus length and response duration with a few sessions of training.

Third, the subjects learned the effects of stimulus angle at a slower rate than they learned the effects of stimulus length, and they continued to use power component functions with respect to angle after three sessions of training. A similar pattern of differential learning rates with respect to length and angle was also found in Experiment 1 and is discussed further in General Discussion.

EXPERIMENT 3

The results of Experiments 1 and 2 speak strongly for the hypothesis that in learning to relate line length and angle to duration, people are initially biased toward inducing a multiplicative combination of power functions. The results of Experiment 2 also suggest that people are quite flexible and can learn other types of multidimensional function. Nonetheless, one might argue that the ad-

ditive combination of linear functions that the subjects learned in Experiment 2 is special in that a logarithmic transformation of stimulus and response variables turns a multiplicative combination of power functions to an additive combination of linear functions. Thus, it is possible that the subjects learned the additive combination of linear functions simply by omitting a logarithmic transformation.

Experiment 3 was designed to investigate how general the subjects' strategy of induction is. This was done by examining how well the subjects could learn a two-dimensional function involving a combination rule somewhat more complex than the ones used in the previous experiments. Experiment 3 also provided a second opportunity to examine whether people are initially biased to induce a multiplicative combination of power functions.

Method

The experimental method paralleled those of Experiments 1 and 2, except that the correct responses were related to the stimuli by a function $D = -733.5 + 18.26L + 31.67A - .2883LA$, where D, L, and A denote the duration (in milliseconds), length (in millimeters), and angle of orientation (in degrees). The stimulus lengths, angles, and response durations are listed in Table 7. The responses spanned about the same range as those in Experiments 1 and 2. The stimulus lengths and angles had narrower ranges than those used in Experiments 1 and 2, but they were still highly discriminable.

Given the relatively narrow ranges of stimulus lengths and angles used, logarithmic transformations of the stimulus and response variables have only a small effect on the form of the function. In both untransformed and logarithmically transformed coordinates, the required function has negligible or no quadratic terms and a large interaction term (see Figure 6 below). In other words, the

Table 7
Stimulus-Response (S-R) Pairs in Experiment 3

S-R Pair	Stimulus Length (mm)	Stimulus Angle (degree)	Response Duration (msec)
1	34.7	20	333
2	34.7	30	550
3	34.7	40	767
4	34.7	50	983
5	34.7	60	1200
6	52.0	20	550
7	52.0	30	717
8	52.0	40	883
9	52.0	50	1050
10	52.0	60	1217
11	69.4	20	767
12	69.4	30	883
13	69.4	40	1000
14	69.4	50	1117
15	69.4	60	1233
16	86.7	20	983
17	86.7	30	1050
18	86.7	40	1117
19	86.7	50	1183
20	86.7	60	1250
21	104.1	20	1200
22	104.1	30	1217
23	104.1	40	1233
24	104.1	50	1250
25	104.1	60	1267

Table 8
Mean Log Response Durations for Each Stimulus-Response (S-R) Pair Averaged Over Subjects as a Function of Half-Session and the Correct Log Response Duration For Each Pair in Experiment 3

S-R Pair	Correct Log Response	Half-Session					
		1	2	3	4	5	6
1	5.808	6.251	6.009	6.041	6.053	5.973	5.979
2	6.310	6.327	6.342	6.284	6.380	6.283	6.306
3	6.643	6.362	6.565	6.583	6.675	6.670	6.648
4	6.891	6.534	6.771	6.761	6.801	6.851	6.845
5	7.090	6.671	6.912	6.921	6.953	6.963	7.007
6	6.310	6.465	6.479	6.453	6.537	6.530	6.545
7	6.575	6.601	6.658	6.719	6.720	6.762	6.801
8	6.783	6.738	6.847	6.822	6.842	6.881	6.858
9	6.957	6.723	6.883	6.888	6.942	6.916	6.940
10	7.104	6.844	6.954	7.010	7.004	7.004	7.018
11	6.643	6.702	6.678	6.789	6.765	6.760	6.758
12	6.783	6.770	6.777	6.872	6.904	6.909	6.910
13	6.908	6.813	6.847	6.946	6.949	7.006	6.964
14	7.018	6.913	6.971	6.976	7.004	7.048	7.025
15	7.117	6.969	7.053	7.080	7.045	7.066	7.076
16	6.891	6.818	6.775	6.876	6.902	6.937	6.927
17	6.957	6.883	6.921	6.976	6.963	7.044	7.009
18	7.018	6.940	6.951	7.059	7.040	7.056	7.051
19	7.076	6.950	7.020	7.042	7.053	7.105	7.056
20	7.131	7.014	7.081	7.150	7.105	7.124	7.100
21	7.090	6.886	6.928	7.027	7.012	7.050	7.036
22	7.104	6.958	7.017	7.048	7.052	7.075	7.111
23	7.117	6.989	7.083	7.099	7.081	7.116	7.077
24	7.131	7.096	7.111	7.149	7.127	7.159	7.094
25	7.144	7.182	7.136	7.222	7.147	7.183	7.110

required combination rule is neither additive nor multiplicative. Thus, the subjects could not learn the required combination rule using a simple strategy of not transforming the stimulus and response variables logarithmically.

Five University of Rochester students participated as paid subjects. Each subject was paid about $5.50 per 1-h session. None had been in Experiment 1 or 2.

Results

Overall stimulus-response relation. Table 8 presents log response durations, averaged over subjects, as a function of stimulus-response pairs and half-sessions. The log response durations for the first half of Session 1 and the last half of Session 3 also appear in Figure 5 as a function of stimulus lengths and angles.

As in Experiment 2, the observed responses systematically deviated from the required ones during the first half-session (top panels). In particular, the observed responses (circles on the solid curves) formed approximately parallel lines, in sharp contrast to the converging dotted curves that represent the required function. Again, it appears that the subjects were biased to induce a multiplicative combination of power functions. The systematic biases, however, decreased over half-sessions. Detailed quantitative analyses support these interpretations.

Multiple regression coefficients. As in Experiments 1 and 2, Equation 2 was fit to the data. For 4 of the 5 subjects, the regression fits were quite good; the percentage of variance accounted for ranged from 95.6% to 97.0%, with an average of 96.2%. For the remaining sub-
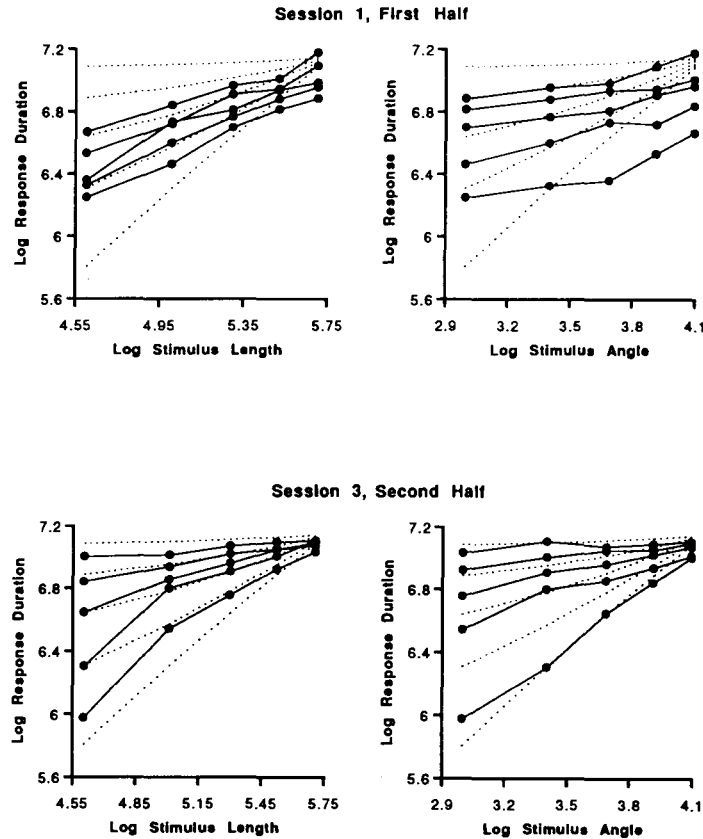
Figure 5. Mean log response durations (natural logarithms of milliseconds) averaged over subjects as a function of log stimulus lengths (natural logarithms of millimeters) and log stimulus angles (natural logarithms of degrees from horizontal) for the first half of Session 1 and the second half of Session 3 of Experiment 3. (The dotted curves represent the required function, and the circles on the solid curves represent the observed mean log response durations.)

ject, the percentage of variance accounted for was considerably lower (86.3%). Nonetheless, ANOVAs performed with and without the last subject's data yielded essentially the same results. The results based on all 5 subjects' data are reported here.

The obtained regression coefficients averaged over subjects appear in Figure 6 as a function of half-sessions. Solid lines represent the ideal values of the coefficients, computed by fitting Equation 2 to the correct log response durations (the regression fit accounts for 99.88% of the variance of the correct log responses). The dotted lines represent zero.

The obtained intercepts (c_1) were initially smaller than the ideal, but increased substantially over half-sessions [$F(5,20) = 4.89, MS_e = .0022, p < .005$]. The log-linear coefficients for length (c_2) were on average not significantly different from the ideal [$F(1,4) < 1, MS_e = .0033$]. As in Experiments 1 and 2, the log-linear coefficients for angle (c_3) were generally smaller than the ideal [$F(1,4) = 7.14, MS_e = .0152, p < .1$]. However, there was a significant improvement over half-sessions [$F(5,20) = 2.88, MS_e = .00065, p < .05$].

Figure 6 shows that the ideal value (solid line) of the coefficient for the interaction between log length and log angle (c_4) is quite different from zero (dotted line). The obtained coefficients during the first half-session were much closer to zero [$F(1,4) = 5.44, MS_e = .00045, p < .1$] than to the ideal value [$F(1,4) = 199.6, MS_e = .00045, p < .001$], but approached the ideal value with practice [$F(5,20) = 25.17, MS_e = .00025, p < .0001$]. A reasonable interpretation of this outcome is that the subjects initially used a multiplicative (log-additive) combination rule, but gradually switched to the required rule (or one similar to it). It is unlikely that the subjects adopted a simple strategy of not transforming the stimulus and response variables logarithmically (i.e., fitting an additive combination of linear functions) when a multiplicative rule failed. The reason is that the expected value of c_4 under the strategy is $-.031$, which is far less negative than the values observed in the later stages of learning.[7]

The ideal values of the log-quadratic coefficients for length and angle (c_5 and c_6) are almost zero (see Figure 6). The obtained coefficients were also close to zero. Mean c_6 was not statistically different from zero [$F(1,4) < 1$,
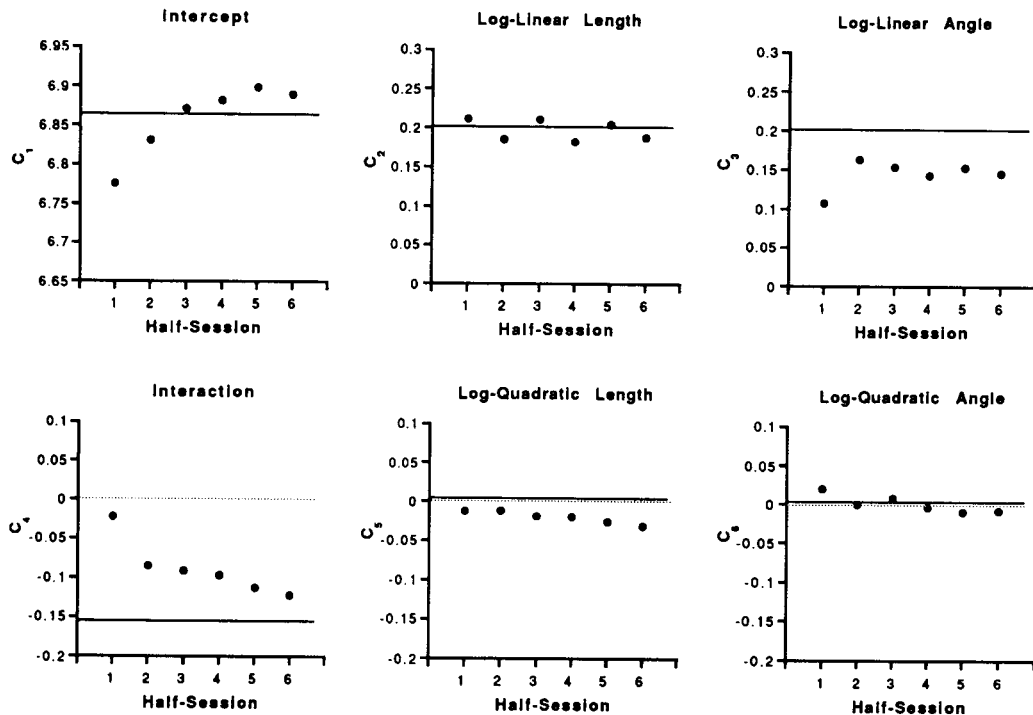
Figure 6. Multiple regression coefficients averaged over subjects as a function of half-sessions in Experiment 3. (The solid lines represent the ideal values of the coefficients, and the dotted lines represent zero.)

$MS_e = .0016$]. Mean c_5, however, was slightly but reliably negative [$F(1,4) = 26.96, MS_e = .00044, p < .01$]. What this unexpected negative log-quadratic trend means is unclear. Yet, the trend is relatively small, and probably does not affect the conclusions drawn from other aspects of the present results.

Variable errors. Table 9 lists the variable errors averaged across subjects for each combination of half-sessions and stimulus-response pairs. As in previous experiments, the variable errors decreased substantially over half-sessions [$F(5,20) = 12.81, MS_e = .023, p < .0001$]. The average variable errors for the six half-sessions were .276, .185, .173, .151, .157, and .146. The variable errors also differed significantly across stimulus-response pairs [$F(24,96) = 6.56, MS_e = .005, p < .0001$]. The variable errors in Experiment 3 were comparable in magnitude to those found in Experiments 1 and 2. Again, it appears that the type of required function had an effect on the pattern of response biases, but not on response variability.

Discussion

The primary outcome of Experiment 3 closely resembled that of Experiment 2. During the early phases of learning, the subjects showed systematic response biases toward a multiplicative combination of power functions, but the biases gradually decreased over the sessions. This outcome provides strong evidence against the possibility that subjects simply leave out the logarithmic transfor-

Table 9
Standard Deviations of Log Response Durations for Each Stimulus-Response (S-R) Pair Averaged Over Subjects as a Function of Half-Session in Experiment 3

S-R Pair	Half-Session					
	1	2	3	4	5	6
1	0.318	0.239	0.262	0.251	0.156	0.182
2	0.345	0.307	0.245	0.187	0.213	0.195
3	0.444	0.224	0.191	0.173	0.166	0.186
4	0.365	0.175	0.188	0.168	0.182	0.179
5	0.363	0.170	0.169	0.140	0.166	0.138
6	0.307	0.275	0.314	0.256	0.214	0.171
7	0.304	0.207	0.186	0.143	0.165	0.164
8	0.226	0.152	0.153	0.125	0.161	0.168
9	0.295	0.172	0.173	0.169	0.164	0.156
10	0.219	0.183	0.135	0.141	0.143	0.151
11	0.258	0.188	0.210	0.166	0.254	0.168
12	0.238	0.183	0.184	0.124	0.161	0.128
13	0.250	0.199	0.144	0.129	0.141	0.124
14	0.211	0.155	0.154	0.129	0.143	0.124
15	0.257	0.194	0.173	0.131	0.128	0.116
16	0.305	0.181	0.211	0.165	0.164	0.168
17	0.277	0.174	0.121	0.146	0.152	0.106
18	0.275	0.147	0.144	0.134	0.162	0.122
19	0.260	0.157	0.144	0.149	0.128	0.117
20	0.286	0.164	0.120	0.137	0.130	0.123
21	0.282	0.215	0.156	0.149	0.139	0.141
22	0.241	0.169	0.143	0.123	0.140	0.119
23	0.132	0.149	0.140	0.120	0.125	0.134
24	0.238	0.116	0.144	0.138	0.121	0.132
25	0.222	0.151	0.135	0.098	0.121	0.141

mation of the stimulus and response variables when the function they must learn deviates significantly from a multiplicative combination of power functions. Instead, it appears that subjects' strategy of induction is general enough for acquiring combination rules other than multiplicative and additive ones.

GENERAL DISCUSSION

Summary of Results

The present research provides a clear demonstration that in some situations people can learn multiplicative combinations of power functions more easily than other two-dimensional functions (e.g., additive combinations of linear functions). In learning to relate stimulus length and angle to response duration, the subjects behaved initially as if they were inducing a multiplicative combination of power functions. This happened regardless of whether the required two-dimensional function was a multiplicative combination of power functions (Experiment 1), an additive combination of linear functions (Experiment 2), or a more complex combination of linear functions (Experiment 3). In particular, systematic response biases occurred during the early phases of learning in Experiments 2 and 3. The biases, however, decreased substantially across sessions.

Relation to Past Studies of Function Learning

The present results replicate and extend Koh and Meyer's (1991) finding that in unidimensional function learning people have a bias toward inducing power functions. On the other hand, the results seem to be in conflict with other previous results showing that people learn additive combination rules and linear functions more easily than nonadditive combination rules and nonlinear functions (Brehmer, 1969, 1974; Carroll, 1963; Deane et al., 1972; Hammond & D. A. Summers, 1965; S. A. Summers et al., 1969). The two sets of results (i.e., those showing the primacy of multiplicative combinations and power functions and those showing the primacy of additive combinations and linear functions), however, are not as different as they might seem at first. As noted earlier, multiplicative combinations of power functions and additive combinations of linear functions are closely related in that a logarithmic transformation of stimulus and response variables turns the former into the latter. Thus, a parsimonious account of both sets of results would be that people have an a priori bias for inducing additive combinations of linear functions, except that, under some circumstances, stimulus and response variables are first transformed logarithmically.

According to this account, one of the main research questions about function learning is the question of what factors determine whether or not people transform the stimulus and response variables logarithmically for a particular function learning task. At what stage and why does the logarithmic transformation occur? The present experiments were not designed to address these issues specifically, but some speculation may be in order.

There are two possible stages at which the logarithmic transformation could occur. One possibility is that certain stimulus and response variables are transformed logarithmically during the initial encoding stage (i.e., their psychophysical functions are logarithmic). Despite arguments made by S. S. Stevens and his colleagues in favor of power psychophysical functions (J. C. Stevens, Mack, & S. S. Stevens, 1960; S. S. Stevens, 1956, 1957, 1961; S. S. Stevens & Guirao, 1963), the existing empirical evidence does not exclude the possibility of logarithmic psychophysical functions (Ekman, 1964; MacKay, 1963; Shepard, 1981). However, the simple hypothesis that some variables are always encoded logarithmically is unlikely to be true, given that line length was used in some of the studies showing the primacy of additive combinations of linear functions (Brehmer, 1969), as well as in the present study. It may be that the way a stimulus variable is encoded depends on the response modality. Perhaps, the psychophysical functions for line length and orientation are linear when numerical estimates or line lengths are used as responses, but are logarithmic when durations are used as responses. This is not unreasonable in light of the findings in the psychophysical scaling literature suggesting that different procedures (e.g., magnitude estimation vs. category judgment) yield different psychophysical functions (S. S. Stevens & Galanter, 1957; S. S. Stevens & Guirao, 1962).

Alternatively, the logarithmic transformation may occur at some later stage, depending on people's (unconscious) hypothesis about the most likely type of function to be learned in a given situation. A multiplicative combination of power functions has a simplest (linear and additive) representation when its variables are transformed logarithmically. Thus, when the required function is expected to be a multiplicative combination of power functions, it would make sense for the learning mechanism to first transform the variables logarithmically so as to simplify the form of the function to be induced. In this way, the same inductive process that has a predisposition toward an additive combination of linear functions can be used to deal with situations in which multiplicative combinations of power functions are likely, as well as situations in which additive combinations of linear functions are likely.

These considerations raise the intriguing question of why multiplicative combinations of power functions might be considered most likely in some situations. (This question is independent of the question of whether the logarithmic transformations occur during early sensory encoding or during later inductive processing.) One reason the subjects in the present experiments exhibited a bias toward inducing a multiplicative combination of power functions may be that physical laws concerning variables such as distance and duration are often multiplicative combinations of power functions.[8] People may have evolved a learning mechanism with a bias toward inducing these functions in order to interact with the physical environment efficiently. The perceptual and motor variables (especially the motoric nature of the keypress responses) used

in the present experiments might have allowed this mechanism to come into play. In contrast, estimating numbers is a cognitive rather than a sensorimotor activity. Perhaps people are biased toward inducing an additive combination of linear functions among numerical estimates because mental addition is practiced more frequently than is mental multiplication.

In summary, the nature of the stimulus and response variables involved may influence the a priori assumptions people make about the type of function to be learned, which in turn determine whether or not logarithmic transformations of the variables (either during sensory encoding or during later inductive processing) occur. Below, these ideas are formalized as an extension of the adaptive regression model of function learning (Koh & Meyer, 1991).

Adaptive Regression Model

The adaptive regression model was proposed by Koh and Meyer (1991) to account for their finding that people learn power functions more readily than other types of function. According to it, people learn unidimensional stimulus–response relations through a process analogous to statistical regression. The model assumes that each trial with feedback leaves a memory trace of the magnitudes of the stimulus and correct feedback response. The stimulus and correct-response magnitudes are first transformed logarithmically, and then a polynomial (e.g., cubic) regression is performed. The logarithmic transformation of the stimulus and response magnitudes turns power functions into linear functions, allowing a simple implementation of the power-function bias as a constraint on the curvature of the fitted polynomial function.

In a standard regression analysis, parameters (polynomial coefficients) are estimated by minimizing a sum of squared deviations between data and the fitted regression function. In the adaptive regression model, however, the parameters are estimated so as to minimize a weighted combination of (1) the sum of squared deviations and (2) the curvature of the fitted function in logarithmic coordinates. This method of parameter estimation is analogous to regularization methods often used in computational vision (Poggio & Girosi, 1990; Poggio, Torre, & Koch, 1985). The curvature component represents the degree to which the fitted function deviates from a power function. The relative contribution of the second component decreases with practice, and the model can gradually learn nonpower functions as well.[9]

The adaptive regression model, as originally formulated, has a few limitations. First, it applies only to unidimensional function learning. Second, it does not allow the possibility that, under some circumstances, linear functions are more easily learned than power functions. The model is therefore extended to account for multidimensional function learning and is modified such that the logarithmic transformations of stimulus and response variables are optional. Although the revised model can be applied to functions involving more than two variables,

for simplicity, I will use two-dimensional functions to describe it. According to the revised model, the response to a stimulus is selected by means of either a polynomial function

$$R = f(S,T) = \sum_{j=1}^{k} a_j S^{bj} T^{cj}, \tag{3a}$$

or a log-polynomial function

$$\log R = f(U,V) = \sum_{j=1}^{k} a_j U^{bj} V^{cj}$$

$$= g(S,T) = \sum_{j=1}^{k} a_j (\log S)^{bj} (\log T)^{cj}, \tag{3b}$$

in which S, T, and R are the magnitudes of the two stimulus variables and the response variable, respectively. U and V denote $\log S$ and $\log T$, respectively. Each a_j represents a polynomial coefficient, and b_j and c_j are nonnegative integers representing the powers to which the stimulus variables are raised. Equation 1, introduced in Experiment 1, is an example of the log-polynomial function.

The coefficients of the function are estimated as in the original adaptive regression model on the basis of the memory traces of the stimuli and corresponding feedback responses. Specifically, the estimation process attempts to minimize the following loss (badness of fit) function:

$$L = \lambda L_1 + (1 - \lambda) L_2, \tag{4}$$

in which L_1 is the sum of squared deviations, L_2 is the curvature of the fitted function, and λ is the weight that controls the relative contributions of the two components $(0 \le \lambda \le 1)$.

The first component L_1, which represents the closeness of the fitted function to the stored data, is defined as

$$L_1 = \sum_{i=1}^{n} [f(S_i,T_i) - F_i]^2, \tag{5a}$$

or

$$L_1 = \sum_{i=1}^{n} [g(S_i,T_i) - \log F_i]^2, \tag{5b}$$

in which S_i, T_i, and F_i denote, respectively, the magnitudes of the two stimulus variables and the feedback response variable of the ith stimulus–response trace; n is the number of stored stimulus–response traces, and $f(S_i,T_i)$ and $g(S_i,T_i)$ are the fitted polynomial and log-polynomial functions defined in Equations 3a and 3b.

The second loss component L_2 is the curvature (i.e., deviation from a plane) of the fitted function defined as

$$L_2 = \int_{Y_l}^{Y_u} \int_{X_l}^{X_u} \left\{ \left[\frac{f^2(x,y)}{\partial x^2} \right]^2 + 2 \left[\frac{f^2(x,y)}{\partial x \partial y} \right]^2 + \left[\frac{f^2(x,y)}{\partial y^2} \right]^2 \right\} \partial x \partial y, \tag{6}$$

in which X_l and Y_l represent the lower bounds of the stim-

ulus ranges (either untransformed or logarithmically transformed), and X_u and Y_u represent their upper bounds; $f^2(x,y)/\partial x^2$, $f^2(x,y)/\partial x\partial y$, and $f^2(x,y)/\partial y^2$ denote the second-order partial derivatives of $f(x,y)$, where f refers to the polynomial function defined in Equations 3a and 3b (Lancaster & Salkauskas, 1986). This curvature component embodies the constraint that treats data as if an additive combination of linear functions (when Equations 3a and 5a are used) or a multiplicative combination of power functions (when Equations 3b and 5b are used) characterizes them. Even if the parameter λ in Equation 4 remains fixed throughout learning, the relative contribution of the curvature component to the loss function decreases as the number of stimulus–response traces increases, thus allowing the fitted function to gradually approach functions other than the preferred type. However, by permitting λ to change over time, a slower or faster rate of convergence may be achieved.[10]

A computer simulation was performed to determine how well the model captures the regularities found in the present data. In the simulation, all four variables involved (two stimulus variables, a feedback variable, and a response variable) were transformed logarithmically—that is, Equation 3b (up to the quadratic terms) and Equation 5b were used. The four variables were treated as normally distributed random variables. Preliminary simulation results were obtained with two free parameters, a common standard deviation (σ) for the four random variables and λ in Equation 4. The same set of parameter values was used to simulate all three experiments. The value of the parameter σ was chosen so that the overall variability for the simulation experiments equaled the average observed variability. The value of the parameter λ was chosen so as to minimize the sum of squared differences between the simulated and actual mean responses. There was a total of 450 mean responses (for three experiments $\times$ six half-sessions $\times$ 25 stimulus–response pairs). Best-fitting parameter values were $\sigma = .164$, and $\lambda = .25$, yielding root mean squared errors of .062, .054, and .073, for the three experiments, respectively.

Even though the fits seemed reasonably good, the preliminary results also revealed a few differences between the model's data and human data. First, the rate of decrease in systematic response biases observed in human data surpassed the rate of decrease obtainable with a fixed value of λ. In order to increase the rate of change in response bias, λ for the ith half-session was set to $i^*\lambda_0$. (An increase in λ corresponds to a decrease in influence of the curvature component.) Second, unlike the human subjects, the model learned the effects of angle and length at the same rate. In order to simulate the difference in learning rate for length and angle, a third parameter ($\alpha > 1$) was added. (More specifically, α for the ith half-session was set equal to $1 + \alpha_0/i$, where $\alpha_0 > 0$.) The parameter α was used to set the standard deviations for length and angle equal to σ/α and $\sigma^*\alpha$, respectively. In other words, α was used to simultaneously decrease the variability along the length dimension and increase the

variability along the angle dimension. The difference in variability along the two stimulus dimensions may be interpreted in terms of a difference of perceptual discriminability. An alternative interpretation is that even though the two stimulus dimensions were equally discriminable, the subjects paid more attention to length than to angle, and that paying more attention to a stimulus dimension is analogous to being able to remember the values along the dimension more accurately. According to this latter interpretation, the parameter α has a similar role as the attentional weight parameter used in Nosofsky's (1986) context model of categorization.

The revised model with three parameters (i.e., σ, λ_0, and α_0) was fit to the data using the same criteria as described above. The fits were considerably better than before. Best-fitting parameter values for the revised model were $\sigma = .164$, $\lambda_0 = .12$, and $\alpha_0 = 1.3$, yielding root mean squared errors of .052, .040, and .063, for the three experiments, respectively. The close fit between the model's data and human subjects' data can be seen in Figure 7, which presents the simulation results for the three experiments along with the corresponding human results. Mean log responses (simulated and actual) are plotted as a function of log stimulus lengths for the first and last half-sessions. The closed symbols on the solid curves represent the actual human data (from Figures 1, 3, and 5), and the open symbols on the dotted curves represent the simulated data. Overall, the revised adaptive regression model accounts for the present results quite well.[11] This is particularly impressive given that the same parameter values were used to simulate all three experiments.

The simulated data reproduced all of the major trends in the actual data. First, in Experiments 2 and 3, both the model and human subjects exhibited systematic response biases toward a multiplicative combination of power functions (i.e., parallel lines) during the first half-session. Second, the response biases decreased substantially with practice. By the last half-session, the responses (both simulated and actual) were much closer to the required responses. Third, the model learned the effects of angle at a slower rate than it learned the effects of length. Multiple regression analyses of the simulated data showed that the coefficients for angle were smaller than the coefficients for length throughout training, and that the angle coefficients remained smaller than the ideal even after three sessions of training. In other words, the difference in variability of the two stimulus variables (or the parameter α) had a lasting effect on the regression coefficients. As pointed out earlier, this happened because measurement errors in independent (or stimulus) variables result in biased regression coefficients (Hays, 1988).

In conclusion, the present research provides strong support for the central idea behind the adaptive regression model—namely, that function learning involves a compromise between evidence from specific examples and general a priori constraints on the form of the function (cf., Poggio & Girosi, 1990). It also raises several important questions: What factors determine which type of

Experiment 1   Experiment 2   Experiment 3

Session 1, First Half
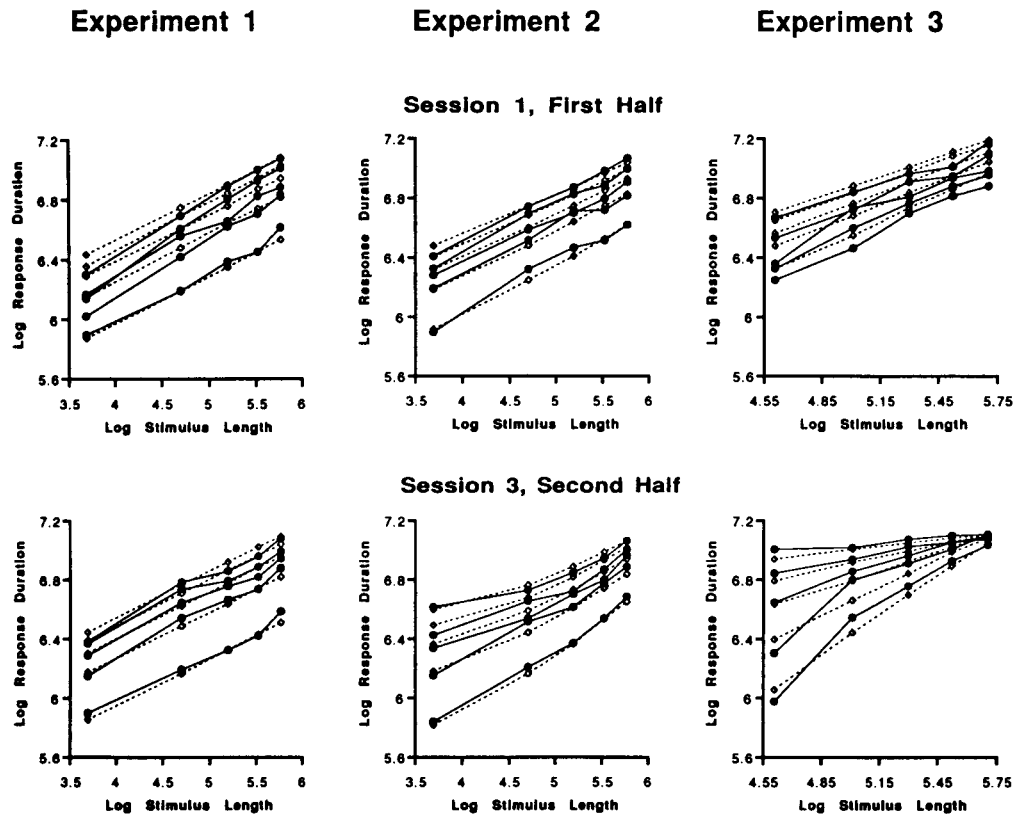

Session 3, Second Half

Figure 7. Simulated and actual data for the first half of Session 1 and the second half of Session 3 of Experiments 1, 2, and 3. (Log response durations are plotted as a function of log stimulus lengths. The closed symbols on the solid curves represent the actual human data, which are the same as those in Figures 1, 3, and 5, and the open symbols on the dotted curves represent the simulated data. The same parameter values were used to simulate all three experiments.)

function is learned most readily in a particular function learning situation? Why are certain aspects of a task learned more slowly than others? The present research paradigm, with its emphasis on analysis of response biases as a function of learning, may prove to be a powerful tool for addressing these questions.

REFERENCES

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Anderson, N. H. (1983). Intuitive physics: Understanding and learning of physical relations. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analyses* (pp. 231-265). Hillsdale, NJ: Erlbaum.

Barsalou, L. (1989). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition, III: Content and process specificity in the effects of prior experiences* (pp. 61-88). Hillsdale, NJ: Erlbaum.

Bjorkman, M. (1965). Studies in predictive behavior. *Scandinavian Journal of Psychology*, 6, 129-156.

Brehmer, B. (1969). Cognitive dependence on additive and configural cue-criterion relations. *American Journal of Psychology*, 82, 490-503.

Brehmer, B. (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior & Human Performance*, 11, 1-27.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, 45, 223-241.

Busemeyer, J. R., Myung, I. J., & McDaniel, M. A. (in press). Cue competition effects in prediction: Experimental tests of adaptive network learning models. *Psychological Science.*

Carroll, J. D. (1963). *Functional learning: The learning of continuous functional maps relating stimulus and response continua* (ETS RB 63-26). Princeton, NJ: Educational Testing Service.

Deane, D. H., Hammond, K. R., & Summers, D. A. (1972). Acquisition and application of knowledge in complex inference tasks. *Journal of Experimental Psychology*, 92, 20-26.

Ekman, G. (1964). Is the power law a special case of Fechner's law? *Perceptual & Motor Skills*, 19, 730.

Hammond, K. R., & Summers, D. A. (1965). Cognitive dependence on linear and nonlinear cues. *Psychological Review*, 72, 215-224.

Hays, W. L. (1988). *Statistics* (4th ed.). New York: Holt, Rinehart, and Winston.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.

Klayman, J. (1988). On the how and why (not) of learning from outcomes. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT approach* (pp. 115-162). Amsterdam: North-Holland.

Koh, K., & Meyer, D. E. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 17, 811-836.

Lancaster, P., & Salkauskas, K. (1986). *Curve and surface fitting: An introduction.* New York: Academic Press.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review*, 66, 81-94.

LUCE, R. D., & TUKEY, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1-27.

MACKAY, D. M. (1963). Psychophysics of perceived intensity: A theoretical basis of Fechner's and Stevens' laws. *Science*, 139, 1213-1216.

MALONEY, L. T., & LANDY, M. S. (1989). Psychophysical estimation of the human depth combination rule. In W. A. Pearlman (Ed.), *Visual communication and image processing: IV. Proceedings of the SPIE* (Vol. 1199, pp. 1154-1163). Bellingham, WA: SPIE.

MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.

NOSOFSKY, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10, 104-114.

NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.

POGGIO, T., & GIROSI, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247, 978-982.

POGGIO, T., TORRE, V., & KOCH, C. (1985). Computational vision and regularization theory. *Nature*, 317, 314-319.

PRICE, P., MEYER, D. E., & KOH, K. (1992). Multivariable function learning: Applications of the adaptive regression model to intuitive physics. In J. K. Kruschke (Ed.), *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 552-557). Hillsdale, NJ: Erlbaum.

SHEPARD, R. N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology*, 24, 21-57.

STEVENS, J. C., MACK, J. D., & STEVENS, S. S. (1960). Growth of sensation on seven continua as measured by force of handgrip. *Journal of Experimental Psychology*, 59, 60-67.

STEVENS, J. C., & MARKS, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Science*, 54, 407-411.

STEVENS, S. S. (1956). The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, 69, 1-25.

STEVENS, S. S. (1957). On the psychophysical law. *Psychological Review*, 64, 153-181.

STEVENS, S. S. (1959). Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology*, 57, 201-209.

STEVENS, S. S. (1961). To honor Fechner and repeal his law. *Science*, 133, 80-86.

STEVENS, S. S. (1965). Matching functions between loudness and ten other continua. *Perception & Psychophysics*, 1, 5-8.

STEVENS, S. S., & GALANTER, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 54, 377-411.

STEVENS, S. S., & GUIRAO, M. (1962). Loudness, reciprocality, and partition scales. *Journal of the Acoustical Society of America*, 34, 1466-1471.

STEVENS, S. S., & GUIRAO, M. (1963). Subjective scaling of length and area and the matching of length to loudness and brightness. *Journal of Experimental Psychology*, 66, 177-186.

SUMMERS, S. A., SUMMERS, R. A., & KARKAU, V. T. (1969). Judgments based on different functional relationships between interacting cues and a criterion. *American Journal of Psychology*, 82, 203-211.

THOMPSON, K. (1990). Rendering anti-aliased lines. In A. S. Glassner (Ed.), *Graphics gems* (pp. 105-106). New York: Academic Press.

TVERSKY, A. (1967). Additivity, utility, and subjective probability. *Journal of Mathematical Psychology*, 4, 175-201.

ULRICH, R., & GIRAY, M. (1989). Time resolution of clocks: Effects on reaction time measurement—Good news for bad clocks. *British Journal of Mathematical & Statistical Psychology*, 42, 1-12.

WALLSTEN, T. S. (1976). Using conjoint-measurement models to investigate a theory about probabilistic information processing. *Journal of Mathematical Psychology*, 14, 144-185.

NOTES

1. S. S. Stevens's (1965) exponents for length and duration are close to 1.0, suggesting that the physical and perceptual magnitudes for these dimensions are proportional to each other. In addition, multidimensional scaling solutions for stimuli similar to the present ones (varying in length and angle of orientation) have indicated that the perceptual space is linearly related to the physical space (Nosofsky, 1986). I therefore assume that the psychophysical transformations for the stimulus and response dimensions are proportional.

2. Compared with a 1-msec resolution timer, a 12-msec resolution timer may seem quite inaccurate. As pointed out by Ulrich and Giray (1989), however, the added uncertainty in timing has only a small effect on the distribution of the measured response durations (i.e., no change in the mean and an increase of a few milliseconds in the standard deviation).

3. If lines are drawn on the AppleColor monitor using only two gray levels (i.e., black and white), they can appear jagged, depending on the orientation. With the full use of the 256 gray levels available on the monitor, however, lines that appear smooth at all orientations can be drawn. The antialiasing algorithm described in Thompson (1990) was used to create smooth lines.

4. The present experimental paradigm is a natural two-dimensional extension of the one developed by Koh and Meyer (1991) for unidimensional function learning. There is, however, one major difference. The studies of Koh and Meyer included some transfer stimulus-response pairs about which subjects did not receive any feedback. The transfer pairs provided valuable data concerning the generalization process, but they were not necessary for assessing relative learnability of functions. The reason is that systematic response biases, which occurred not only for transfer pairs but also for practice pairs, served as a sensitive measure of learnability. Because the main objective of the present experiments concerned learnability of various combination rules, no transfer pairs were included. Nonetheless, trials without feedback were included. This was done to keep the procedure close to the one used by Koh and Meyer (1991). It seems reasonable to assume that the conclusions drawn from the results of the present experiments do not depend on this manipulation.

5. The formula used to compute the point score (P) was $P = \max\{0, [10 - 0.033 \mid D_a - D_c\mid]\}$, where D_a and D_c denote the actual and correct response durations (in milliseconds), respectively.

6. Similar analyses were also performed on untransformed data. A quadratic equation based on the raw stimulus and response variables was formulated as follows:

$$M(D) = b_1 + b_2 L + b_3 A + b_4 LA + b_5 L^2 + b_6 A^2,$$

where $M(D)$ denotes the mean of untransformed response durations, and L and A denote untransformed length and angle. The above equation was orthogonalized and fit to the data. The equation reduces to an additive combination of linear functions when b_4, b_5, and b_6 equal zero. A nonzero value of b_4 indicates a deviation from an additive combination rule, and nonzero values of b_5 and b_6 indicate deviations from linear functions. The conclusions based on these analyses were essentially the same as those based on the analyses of logarithmically transformed data. Analyses of transformed data are reported here because the primacy of multiplicative combinations of power functions can be seen most clearly in logarithmic coordinates.

7. To compute the expected value of c_4 under the strategy of fitting an additive combination of linear functions, the correct response durations were first fit to a regression equation of the form $D = d_1 + d_2 L + d_3 A$, in which D, L, and A denote response duration, stimulus length, and stimulus angle, and d_1, d_2, and d_3 denote regression coefficients. Next, the fitted equation was used to compute the predicted durations. Then, the predicted durations were logarithmically transformed and fit to Equation 2.

8. Luce (1959) showed that when variables lie on ratio scales, power functions (and by extension multiplicative combinations) are the only ones that have the desirable property of invariance under admissible re-

scalings of variables. Not all physical laws are strictly multiplicative combinations of power functions, however. For example, the amount of time (D) that a ball takes to roll down an inclined plane of length L and angle A has the functional form $D = k(L/\sin A)^{.5}$, where k is a constant. Specifically, the angle term involves a sinusoidal transformation. It is possible, as suggested by Price, Meyer, and Koh (1992), that in certain situations people initially assume a linear relationship between logarithmically transformed duration and sinusoidally transformed angle. The present results, however, do not distinguish between this possibility and the alternative possibility that people initially assume a linear relationship between logarithmically transformed duration and logarithmically transformed angle. This is because, for the range of angles used in the present studies, the logarithmic and sinusoidal transformations are not very different. Further experiments with a wider range of angles in a variety of context might prove valuable.

9. Brehmer (1974) proposed a related model of function learning, which is based on the assumption that people have a hierarchically organized set of hypotheses about function types. The difference between Brehmer's model and the adaptive regression model is discussed in detail in Koh and Meyer (1991).

10. The proposed model belongs to the class of rule-based models of inductive learning. It may be possible to formulate an exemplar-based model (analogous to the models developed by Hintzman, 1986, Medin & Schaffer, 1981, and Nosofsky, 1984, 1986, for category learning) that could also account for the present results. However, such an exemplar model would probably include a special mechanism for biasing initial responses toward a certain type of function, as well as a mechanism for reducing the response biases over time. In other words, the exemplar model would be able to account for the present results precisely because it embodies the same computational objective as the proposed rule-based model, and it would be virtually impossible to distinguish between the two models on empirical grounds (Barsalou, 1989). The rule-based approach is chosen here because it allows a simple, straightforward specification of a priori biases.

11. The model showed little decrease in variability over half-sessions. Additional learning mechanisms, such as learning to produce intended response durations accurately, may be necessary to model the observed pattern of decrease in response variability as a function of practice.