Multivariable Function Learning: Applications of the Adaptive Regression Model to Intuitive Physics

Paul C. Price and David E. Meyer
Department of Psychology
University of Michigan
330 Packard Road
Ann Arbor, MI 48104-2994
paul_price@um.cc.umich.edu
david_meyer@um.cc.umich.edu

Kyunghee Koh
Department of Psychology
University of Rochester
Rochester, NY 14627
koh@cvs.rochester.edu

Abstract

We investigated multivariable function learning--the acquisition of quantitative mappings between multiple continuous stimulus dimensions and a single continuous response dimension. Our subjects learned to predict amounts of time that a ball takes to roll down inclined planes varying in length and angle of inclination. Performance with respect to the length of the plane was quite good, even very early in learning. On the other hand, performance with respect to the angle of the plane was systematically biased early in learning, but eventually became quite good. An extention of Koh and Meyer's (1991) adaptive regression model accounts well for the results. Implications for the study of intuitive physics more generally are discussed.

Introduction

Our research concerns function learning--the acquisition of quantitative mappings between continuous stimulus and response dimensions. In tennis, for example, your distance from the net is a continuous stimulus dimension, and the force with which you should hit the tennis ball is a continuous response dimension. Furthermore, there is a function that relates your distance from the net to the appropriate amount of force with which you should hit the ball. You must learn this function if you are to become a competent tennis player.

Of course, in reality optimal response magnitudes are often functions of more than one stimulus dimension. The amount of force with which you should hit a tennis ball depends on more than just your distance from the net. It is also, for example, a function of the velocity with which the ball is approaching you. The present article, therefore, is about multivariable function learning--the acquisition of quantitative mappings between multiple continuous stimulus dimensions and a continuous response dimension.

Before presenting our results on multivariable function learning, we will review some recent work by Koh and Meyer (1989, 1991) on single-variable function learning. The work by Koh and Meyer--particularly their adaptive regression model of function learning--is highly relevant here because it forms the basis for our current research on multivariable function learning.

The Work of Koh and Meyer (1991)

Koh and Meyer (1991) studied the learning of a motor response to a single dimension of a perceptual stimulus. On each trial of their experiments, subjects were presented with a stimulus consisting of two vertical lines separated by some horizontal distance. Given this distance, subjects had to make two successive finger taps such that the amount of time between the taps equalled a correct response duration, which was a predetermined function of the stimulus length. There were two types of stimulus-response pairs: practice and test. At the end of each trial with a practice stimulus-response pair, subjects received auditory feedback consisting of two beeps, with the amount of time between the beeps equalling the correct response duration for the current stimulus. Subjects were also informed about whether their response durations had been too long or too short, and given a point score between 0 and 100, depending on how close their actual response duration had come to the correct response duration. Subjects received no feedback after trials with test stimulus-response pairs.

In each of three separate experiments performed by Koh and Meyer (1991), the correct response duration was related to the stimulus length by one of three strictly monotone mappings: a power function with a positive exponent less than one, a logarithmic function, or a linear function with a positive intercept. Koh and Meyer (1991) found that regardless of which function had to be learned, subjects' initial responses appeared to be a power function of the stimulus length. This resulted in the power function of Experiment 1 being learned quickly and accurately, whereas the logaritmic and linear functions of Experiments 2 and 3 were learned only after considerable practice.

This power-function bias is explained by Koh and Meyer (1991) with an adaptive regression model of function learning. According to the adaptive regression model, the magnitudes of the stimulus lengths and feedback about correct response durations are transformed logarithmically and stored--along with some noise--in a procedural memory. After each learning trial, a polynomial regression of the transformed responses onto the transformed stimuli is performed. Subsequent responses are then chosen on the basis of parameters derived from this regression.

Under the adaptive regression model, the regression function is initially constrained to be linear in log-log coordinates at the start of learning. This constraint allows people to learn power functions quickly and accurately, because such functions are linear in log-log coordinates. Because other (non-power) functions are not linear in log-log coordinates, the initial linearity constraint imposed by the adaptive regression model implies that they would be learned less easily, consistent with Koh and Meyer's (1991) results. As practice progresses, however, the adaptive regression model assumes that the initial linearity constraint is gradually relaxed, eventually allowing various non-power functions to be learned accurately, just as Koh and Meyer also found.

Mathematically, the adaptive regression model is embodied in the four equations shown below.

(1) $$\ln R = \sum_{j=0}^{k} a_j (\ln S)^j.$$

(2) $$L = \lambda L_1 + (1 - \lambda) L_2.$$

(3) $$L_1 = \sum_{i=1}^{n} \left\{ \left[\sum_{j=0}^{k} a_j (\ln S_i)^j \right] - \ln R_i \right\}^2.$$

(4) $$L_2 = \int_{S_S}^{S_L} \left[\sum_{j=2}^{k} j (j-1) a_j x^{j-2} \right]^2 dx.$$

Here responses are chosen according to Equation 1, where R is the current response duration to be produced and S is the current stimulus length. The coefficients of Equation 1, the a_js, are estimated through the regression process so as to minimize the quantity L of Equation 2. As Equation 2 shows, L is a weighted combination of components L_1 and L_2, whose values appear in Equations 3 and 4. L_1 is simply the sum of squared deviations between predicted response durations and stored feedback about values of correct response durations, and L_2 represents the degree of curvature of the fitted function.

Including L_2 as part of minimizing L biases the computed coefficients of the polynomial's nonlinear terms to have small absolute values whenever the weight parameter λ (which takes a value between 0 and 1) is much less than 1. Consequently, the regression algorithm will tend initially to yield a linear function in log-log coordinates. Nevertheless, if λ is significantly greater than zero, placing some weight on L_1, the adaptive regression model can overcome its initial tendencies. As more and more stimulus-response pairs are experienced, the sum of squared deviations in L_1 will increase, eventually overshadowing the curvature constraint L_2, which remains essentially constant. With an appropriate value of λ, therefore, the model can closely mimic the rate at which people learn non-power functions. (See Koh & Meyer, 1991, for more details).

The Inclined-Plane Experiment

To study multivariable function learning and to extend the adaptive regression model, we had subjects learn to predict the amount of time that a ball takes to roll from the top to the bottom of an inclined plane, based on the length of the plane and its angle of inclination. Our stimuli were presented on a video display, and consisted of line segments varying in length and angle of inclination. Responses consisted of two key taps such that the amount of time between the taps constituted a prediction about the motion time of a ball on the displayed inclined plane. This methodology is advantageous because it is procedurally very similar to the work of Koh and Meyer (1991), and conceptually very similar to prior work with the inclined-plane task (e.g., Anderson, 1983; Bjorkman, 1965).

Method

Subjects. Three University of Michigan students participated in the experiment. They were paid a base wage of $5.00 per 75-minute session, plus a performance bonus described below.

Design. Subjects learned the function

(5) $$T = k [L / \sin A]^{1/2},$$

Stimulus Angle	Stimulus Length 11.96	18.47	28.47	43.97	67.83
80.00	300	372	463	575	714
39.66	372	463	575	714	887
24.43	463	575	714	887	1102
15.54	575	714	887	1102	1370
10.00	714	887	1102	1370	1701

Table 1. *Correct-Response Duration (ms) as a Function of Stimulus Length (mm) and Stimulus Angle (degrees from horizontal).*

where T is the time that a ball takes to roll from the top to the bottom of an inclined plane, k is a constant of gravitation, L is the length of the plane, and A is its angle of inclination. The stimuli were line segments anchored in the lower left corner of a Hewlett-Packard 1437a graphics display and extending diagonally upward to the right. Each of five stimulus lengths was combined with each of five stimulus angles to form 25 unique stimuli. The constant k was chosen so that the correct response times ranged from 300 to 1701 ms (see Table 1).

The subjects participated in one 75-minute experimental session per day over five consecutive days. Each session began with four warm-up trials; responses to the warm-up trials were highly variable and were therefore not analyzed. The remainder of each session was divided into 30 blocks of 25 trials each. Each of the 25 stimuli was presented once per block, in random order.

Procedure. At the start of each trial, one of the 25 stimuli was presented on the display screen. The subject responded by tapping the slash ("/") key twice so that the amount of time between the two taps was his or her prediction about the amount of time that a ball would take to roll down the displayed inclined plane. After responding, subjects received three forms of correct-response feedback. At 300 ms after the second tap, the first of two short beeps (15 ms, 1000 Hz tones) occurred; a second beep followed. The time between the two beeps was the correct response time. The stimulus display was cleared at the onset of the second beep, and the subject was presented with additional information about the accuracy of his or her response. One of three messages--"LONG," "SHORT," or "PERFECT"--was presented, depending on whether the response was greater than, less than, or equal to the correct response duration. Accompanying this message was a numerical point score, ranging from 0 to 100, which indicated how close the subject's response duration had been to the correct duration. This point score was calculated according to the equation $P = 100 - .2 \mid T_C - T_S \mid$, where P is the number of points, T_c is the correct response duration, and T_s is the duration produced by the subject. If P happened to be negative, a score of zero points was awarded. The message and point score were visible for 700 ms and were followed by a 500 ms intertrial interval.

After each trial block, the subject was presented with his or her point total for the block and cumulative point total for the session. Ultimately, the cumulative point total was converted to a bonus payment of $.05 per 1000 points earned. In general, this resulted in a bonus of between $2.00 and $2.50 per session.

Results

Tables 2a and 2b show subjects' mean response durations for each inclined plane during sessions 1 and 5, respectively. To make the patterns in these tables clearer, Figure 1 shows subjects' mean response durations for sessions 1 and 5 versus the log of the stimulus length, averaged across the five stimulus angles. Response durations have been transformed

Stimulus Angle	Stimulus Length 11.96	18.47	28.47	43.97	67.83
80.00	324	451	498	557	597
39.66	434	567	630	734	794
24.43	534	653	773	879	915
15.54	618	833	945	1024	1132
10.00	741	1032	1171	1328	1475

Table 2a. *Subjects' Mean Response Duration (ms) as a Function of Stimulus Length (mm) and Stimulus Angle (degrees from horizontal). Session 1.*

Stimulus Angle	Stimulus Length 11.96	18.47	28.47	43.97	67.83
80.00	317	395	483	578	651
39.66	383	463	596	730	870
24.43	464	577	743	865	1010
15.54	557	731	893	1010	1196
10.00	744	948	1175	1277	1596

Table 2b. *Subjects' Mean Response Duration (ms) as a Function of Stimulus Length (mm) and Stimulus Angle (degrees from horizontal). Session 5.*

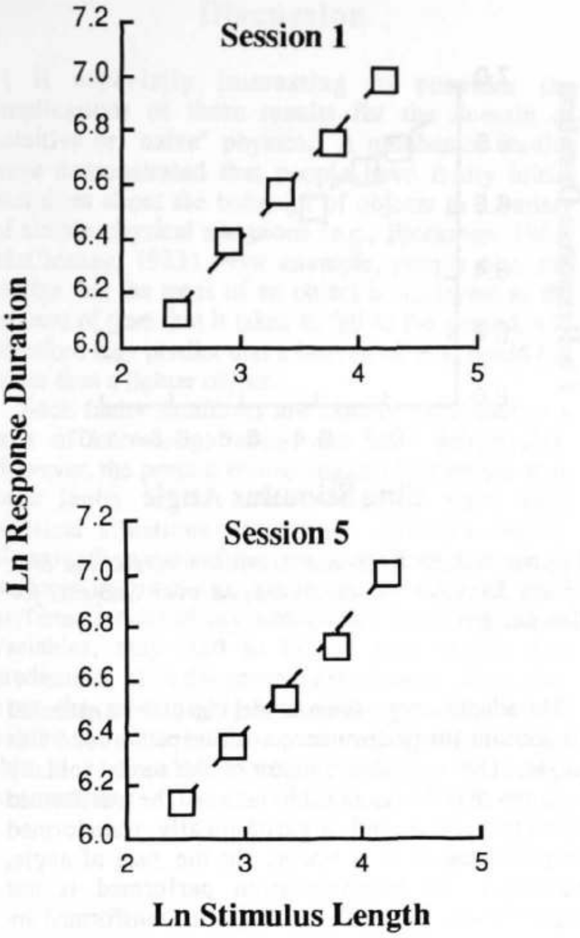

Figure 1. *Log Response Duration versus Log Stimulus Length, Averaged over Subjects and Stimulus Angles for Sessions 1 and 5, respectively.*

logarithmically before averaging the data and making the graphs, so as to produce a linear correct-response function (the dashed lines in Figure 1).

Note that even during session 1, subjects' mean response durations appear to be a nearly log-linear function of stimulus length; that is, they learned this aspect of the inclined-plane task very quickly and accurately. To confirm this, a log-polynomial regression of mean response durations onto stimulus length was performed separately for each subject, session, and angle. Both linear and quadratic coefficients were obtained. The mean value of the linear coefficients was .48, which is not significantly different from the optimal linear coefficient of .50 [$t(2) = 1.54$, $p > .05$]. (Note: The optimal linear coefficient equals 0.50 because of the square-root exponent in Equation 5.) The mean value of the quadratic coefficients was -.00003, which is not significantly different from the optimal quadratic coefficient of zero [$t(2) = .03$, $p > .05$]. This implies that there was essentially no bias in subjects' responses with respect to length even during session 1.

The linear coefficients, or slopes, were then treated as the dependent variable in an ANOVA with session number and stimulus angle as fixed factors and subjects as a random factor. These coefficients did not change significantly across sessions [$F(4,8) = .72$, $p > .05$], nor did they differ significantly across stimulus angles [$F(4,8) = .71$, $p > .05$]. In other words, no significant learning took place with respect to stimulus length after session 1, and performance with respect to stimulus length did not depend on the angle of the stimulus.

Figure 2 shows subjects' mean log response durations for sessions 1 and 5 versus the log of the reciprocal of the sine of the stimulus angle. Again, this produces a linear correct-response function (the dashed lines in Figure 2).

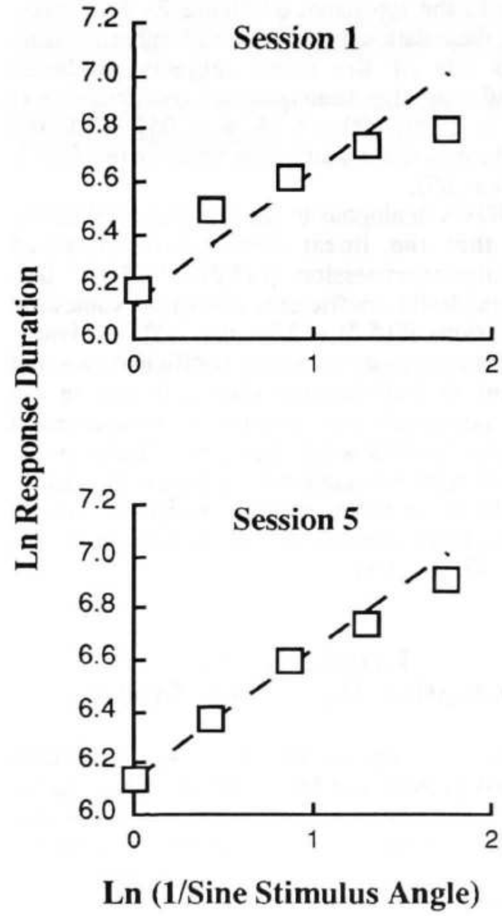

Figure 2. *Log Response Duration versus the Log of the Reciprocal of the Sine of the Stimulus Angle, Averaged over Subjects and Stimulus Lengths for Sessions 1 and 5, respectively.*

Note that during session 1, there is a pronounced curvature in the plotted response times. To confirm this, a log-polynomial regression of mean response duration onto the reciprocal of the sine of the angle was performed for each subject, session, and stimulus length. Both linear and quadratic coefficients were obtained. The mean value of the linear coefficients was .34, which again is not significantly different from the optimal linear coefficient of .50 [$t(2) = 1.80$, $p > .05$]. However, when the data were analyzed separately for each subject, we found that each subject produced linear coefficients that were significantly less than the optimal linear coefficient of .50 [$t(4) = 15.54$, $p < .05$; $t(4) = 12.46$, $p < .05$; $t(4) = 13.16$, $p < .05$].

The mean value of the quadratic coefficients was -.013, which is not significantly different from zero [$t(2) = 1.64$, $p > .05$]. However, it is three orders of magnitude larger than the mean quadratic coefficient for length, and it is easily perceived as the downward curvature in the top panel of Figure 2. Moreover, analyzing these data separately for each subject reveals that two out of the three subjects produced significantly negative mean quadratic coefficients [$t(4) = 13.26$, $p < .05$; $t(4) = 8.26$, $p < .05$], while the third produced a non-significantly negative one [$t(4) = .00005$, $p > .05$].

An ANOVA analogous to the one described earlier showed that the linear coefficients increased significantly across sessions [$F(4,8) = 5.15$, $p < .05$], and the quadratic coefficients decreased somewhat across sessions [$F(4,8) = 2.34$, $p < .15$]. Although the latter change in the quadratic coefficients was not significant at conventional levels, it was in the direction predicted by the adaptive regression model (see below). In other words, subjects' initial response biases with respect to angle did decrease with practice. Neither the linear nor quadratic coefficients differed across stimulus lengths [$F(4,8) = 1.46$, $p > .05$; $F(4,8) = .37$, $p > .05$].

Extending the Adaptive Regression Model

The adaptive regression model--as originally formulated by Koh and Meyer (1991)--quite readily explains performance with respect to stimulus length in the present experiment. According to this model, the relationship between the logarithmically transformed stimulus length and response duration is assumed to be linear. Because the relationship between the length of an inclined plane and a ball's rolling time is in fact linear in log-log coordinates, the model predicts that it should be learned quite rapidly. This is, of course, exactly what was found in the present experiment (Figure 1).

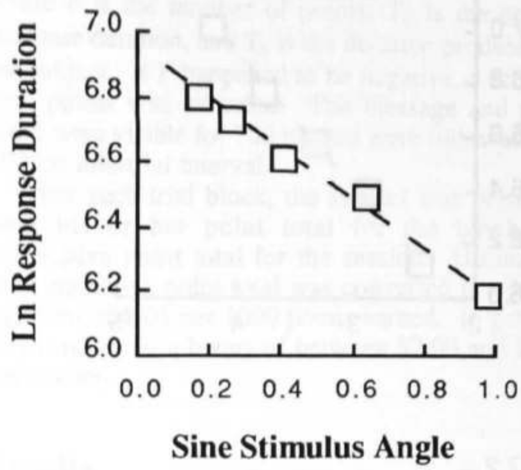

Figure 3. *Log Response Duration versus the Sine of the Stimulus Angle, Averaged over Subjects for Session 1.*

The adaptive regression model can also be extended to account for performance with respect to stimulus angle. Our extended version of the model initially assumes that the relationship between the transformed stimulus angle and logarithmically transformed response duration is linear. In the case of angle, however, the transformation performed is not logarithmic. Instead, the angle is transformed by taking its sine. That subjects actually do assume an initial linear relationship between the log of the response duration and the sine of the angle can be seen in Figure 3. This figure shows subjects' mean log response durations versus the sine of the stimulus angle; the relationship is clearly linear ($r^2 = .994$).

Our extended adaptive regression model assumes further that the transformed response duration is an additive combination of the two transformed stimulus magnitudes: ln L and sin A. Early performance in the inclined-plane task, therefore, is characterized by the equation

$$\ln R = a + b \ln L + c \sin A, \quad (6)$$

where R is the subject's response duration, L is the length of the inclined plane stimulus, and A is its angle of inclination. Equation 6 accounts for 99.4% of the variance in subjects' mean response durations during session 1. Finally, to explain how subjects' performance with respect to stimulus angle improves over sessions, our extended adaptive regression model gradually adds non-zero, higher-order polynomial terms for sine A into Equation 6 as practice progresses, using the same sort of relaxation process posited in Equations 1 through 4.

Discussion

It is especially interesting to consider the implications of these results for the domain of intuitive or "naive" physics. A number of studies have demonstrated that people have faulty initial intuitions about the behavior of objects in a variety of simple physical situations (e.g., Bjorkman, 1965; McCloskey, 1983). For example, people may not realize that the mass of an object is irrelevant to the amount of time that it takes to fall to the ground, and therefore may predict that a heavier object should fall faster than a lighter object.

Such faulty intuitions are usually attributed to a lack of knowledge about the laws of physics. However, the present results suggest that people may have faulty initial intuitions about many such physical situations for a very different reason. Namely, the nature of the transformations that people perform on stimulus variables, as well as their preferred modes of psychologically combining those variables, may lead to biased expectations and predictions, as in the present experiment. If so, then even the most highly educated physicists would exhibit the same biases as naive subjects. Perhaps this is a possibility worth exploring in future experiments.

References

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

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

Koh, K., & Meyer, D.E. (1989). Induction of continuous stimulus-response relations. In G.M. Olson & E.E. Smith (Eds.), *Proceedings of the 11th annual conference of the Cognitive Science Society* (pp. 333-340). Hillsdale, NJ: Erlbaum.

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

McCloskey, M. (1983). Intuitive physics. *Scientific American, 248,* 122-130.