8

Discrimination among two- and three-element sets of weights

ELIZABETH F. SHIPLEY and R. DUNCAN LUCE,
University of Pennsylvania

The purpose of this experiment is to test a choice-theory model for human discrimination. The arguments leading to the model can be found in Luce (1959, pp. 28–34).

Several stimuli—weights, in this experiment—are available, and various subsets are presented to the subject, who is instructed to judge which weight in each presentation is heaviest. He identifies his choice by its position in the order in which he hefted the weights. Thus, if there are three weights, he has three possible responses: first, second, and third. Given a particular order of presentation, the model states that nonnegative numerical scale values can be assigned to each of the possible responses such that the probability of a given response is simply the scale value for that response divided by the sum of the scale values for all the possible responses. Furthermore, the model says that each of these scale values is the product of two other nonnegative numbers, one associated with the response itself and the other with the stimulus designated by that response. Suppose that a given presentation places weight W in the rth response position. Then one number, a ***response-bias parameter***, is assigned to response r independent of the presentation. The other number, which multiplies the first, is associated with W; it depends neither upon the other stimuli in the presentation set nor upon the response that happens to be associated with W in this particular presentation (i.e., its serial position). These numbers we call *stimulus parameters*.

This experiment was conducted in the Laboratory of Social Relations, Harvard University, and was supported in part by National Science Foundation grants NSF G-4506 and G-5544 to Harvard University. The data analysis and the writing, which were mainly carried out at the University of Pennsylvania, were supported in part by National Science Foundation grants NSF G-8864 and NSF G-17637 to the University of Pennsylvania.

We are indebted to several people who have contributed to this study: to Professor Eugene Galanter for advice about the design; to Mr. Curt Fey and Mr. Thomas Angell for preparing computer programs and for supervising computer calculations; to the Computer Center of the University of Pennsylvania, where some of the calculations were carried out; and to Dr. John van Laer for telling us of his iteration scheme for estimating parameters.

A specific example may be helpful. Suppose that we have three weights, H, M, and L (for heavy, medium, and light); let the corresponding stimulus parameters be denoted, respectively, by α, β, and γ, and let the three response-bias parameters be b_1, b_2, and b_3, where all parameters are nonnegative. Then, if the weights are presented in the order H, L, M, the model assigns the scale values αb_1, γb_2, and βb_3 to the three responses, first, second, and third; and the corresponding three response probabilities are

$$\frac{\alpha b_1}{\alpha b_1 + \gamma b_2 + \beta b_3}, \quad \frac{\gamma b_2}{\alpha b_1 + \gamma b_2 + \beta b_3}, \quad \frac{\beta b_3}{\alpha b_1 + \gamma b_2 + \beta b_3}.$$

Note that these probability expressions are unaffected if every term in them is multiplied by the same positive number—this is, of course, equivalent to saying that the scale values lie on a ratio scale. In other words, we are free to select one of the positive weight parameters and one of the positive response-bias parameters to be unity. Our choice will be $\gamma = 1$ and $b_1 = 1$.

Using this notation, we give in Table 1 the scale values for two of the conditions used in the experiment. Although we shall assume that the stimulus parameters remain the same when we vary either the order of presentation or the number of stimuli presented, we have no reason to expect the response

TABLE 1

THE SCALE-VALUE MODEL FOR JUDGMENTS BETWEEN PAIRS (CONDITION 1) AND AMONG TRIPLES (CONDITION 2) OF WEIGHTS

Condition 1			Condition 2			
Stimulus presentations	Response		Stimulus presentations	Response		
	1	2		1	2	3
H,M	αc_1	βc_2	H,M,L	αb_1	βb_2	b_3
H,L	αc_1	c_2	H,L,M	αb_1	b_2	βb_3
M,H	βc_1	αc_2	M,H,L	βb_1	αb_2	b_3
L,H	c_1	αc_2	L,H,M	b_1	αb_2	βb_3
M,L	βc_1	c_2	M,L,H	βb_1	b_2	αb_3
L,M	c_1	βc_2	L,M,H	b_1	βb_2	αb_3
M,M[a]	c_1	c_2	M,M,M[a]	b_1	b_2	b_3

[a] In these presentations, the parameter β is common to all of the responses and therefore has been divided out.

biases to be constant when the number of responses is changed.[1] Therefore a different notation is used for the biases in the two conditions.

For Condition 1, in which pairs of weights are presented, there are seven independent probabilities (the sum of the probabilities in each row in each section of Table 1 must add up to 1) to be accounted for by the three parameters α, β, and c_2. Thus, there are four degrees of freedom if the data are used to estimate the parameters, and a quantity such as χ^2 can be used to evaluate the goodness of fit of the model. For Condition 2, with three weights in each presentation, 14 independent probabilities are to be accounted for by the four parameters α, β, b_2, and b_3, yielding 10 degrees of freedom if the data are used to estimate these parameters. Because we assume that the stimulus parameters are the same from condition to condition, a second test of the model is the similarity of their numerical values. Unfortunately, appropriate statistical tests do not seem to have been worked out, nor does it appear easy to develop them.

1. General method

Subjects. The subjects were three female and six male students, all of whom were naïve about psychological experiments and about the purpose of this experiment. They ranged in age from 18 to 25 years. They were paid an hourly wage plus a bonus that depended upon the accuracy of their judgments.

Conditions. Seven experimental conditions were explored. Each subject participated in either four or six of them (the different conditions employed reflect changes in our thinking; the reasons for these changes are indicated below). Conditions 1 and 2 are stated in Table 1. Conditions 1* and 2* differ from conditions 1 and 2 in two respects: the weights employed were *M*, *L*, and *VL*, where *VL* is lighter than *L*; however, the weight *M* was still used in the *M*,*M* and *M*,*M*,*M* presentations. Thus, in the model for the starred conditions, *M* plays the role of *H* in the unstarred conditions, *L* the role of *M*, and *VL* the role of *L*, except for the presentation of three identical weights. The other three conditions involve exactly the same stimulus presentations as Condition 2; however, subjects were instructed to respond only to two of the three weights. These conditions may be conveniently denoted by (1,2) when the subject is asked to report only about the first two weights, (1,3) when he is asked to report about the first and third, and (2,3) when he is asked to report about the last two. The model corresponding to these three conditions is exactly the same as that for Condition 2 except that the response-bias parameter is 0 in the omitted response column.

Subjects 1 and 2 were run on conditions 2, (1,2), (1,3), and (2,3) with the *M*,*M*,*M* presentations omitted. A block of 102 trials was run on each condition during each daily session. Both the order of presentation of the blocks and the order of presentation of the 102 trials within each block were randomized independently within sessions. A total of 24 experimental sessions were run.

[1] The work of Tanner and his colleagues (for surveys of this work, see Green, 1960; Licklider, 1959; and Luce, 1963) in detection situations strongly suggests that the response biases are under the control of the subject and that they can be manipulated by instructions, by payoffs, and by presentation probabilities.

Subjects 3, 4, and 5 were run on conditions 1, 2, 2*, (1,2), (1,3), and (2,3). The blocks were reduced to 70 trials, but the total number of sessions was increased to 40. The shorter blocks were needed so that the experimental sessions would not exceed two hours.

Subjects 6 through 9 were run on conditions 1, 2, 1*, and 2* in blocks of 105 trials per condition per session, again with randomized presentation of both blocks and trials. A total of 27 experimental sessions were run.

All presentations of either two or three weights within a condition occurred equally often in each session, for a total of at least 400 presentations during the experiment.

Stimuli. The weights were brass cylinders 2 inches high and $\frac{3}{4}$ inches in diameter. They were superficially the same; however, as we shall see, there is some suggestion that at least some subjects were able to identify some of the cylinders, presumably by slight differences in the surfaces. The values used for the several subjects are shown in Table 2.

TABLE 2
WEIGHTS IN GRAMS

Subjects	*H*	*M*	*L*	*VL*
1	103.4	100.0	97.0	—[a]
2	103.0	100.0	97.0	—[a]
3	72.1	70.0	67.9	65.8
4	133.5	130.0	126.5	123.0
5	87.5	85.0	82.5	80.0
6–9	103.0	100.0	97.0	94.0

[a] Not used.

Procedure. Prior to his first run, each subject was given a complete description of the conditions in which he would participate, except that conditions 1 and 2 were not distinguished as being different from their starred counterparts. In other words, subjects were told that three, rather than four, weights were involved. Questioning at the end of the experiment indicated that the subjects believed this. The following payoff procedure was also explained. For correctly naming the heavier of two weights, or of two out of three, the subject received one cent; for naming the lighter as heavier, he lost one cent. For three responses, he received one cent when he selected the heaviest weight, lost one cent when he selected the lightest, and neither lost nor gained when he selected the medium one. Whenever the presentations were the same, no exchange occurred.

Each subject was run individually by a different experimenter for a maximum of two hours a day, at the same time each day. The subject was blindfolded and wore a cotton glove. With his forearm resting upon a support, he lifted the weights in his fingers by swinging his hand from the wrist in time to a metronome set at 92 beats per minute. The first weight of each presentation was lifted on one beat and put down on the next; then the empty hand was moved up and down on the next beats and the second weight was lifted and returned on the fifth and sixth beats. The routine continued if three weights were used. There were four beats, or about 2.60 seconds, between successive lifts in a presentation. After lifting the two or three weights of a given trial, the subject reported which he thought was the heaviest by saying "One," "Two," or "Three." The experimenter then reported the actual order in which the weights were presented. For example, when M,H,L was presented, he said "Medium, heavy, light." When the set M,L,VL was used, he said, "Heavy, medium, light." The time that elapsed between trials in each block (i.e., from putting down the last weight of one trial to lifting the first weight of the next) was seven beats, or about 4.55 seconds.

After each block of trials, the subject rested and the experimenter calculated and announced the total amount earned. Wages and bonuses were paid to the subject after each session.

The experimental sessions were preceded by five training sessions with trials exactly the same as the experimental ones; they are not included in the following analyses.

2. Parameter estimation

Given experimental estimates of the choice probabilities, it is necessary to estimate model parameters for each condition: α, β, b_2 and b_3, or c_2 for conditions 1 and 2; and β, δ (corresponding to the VL weight), b_2 and b_3, or c_2 for the starred conditions. One would like an "optimal" procedure such as maximum likelihood or minimum χ^2, but the postulated nonlinear relation between the probabilities and the scale values leads to systems of nonlinear simultaneous equations that have not been solved. Thus, some sort of iterative numerical procedure had to be used. To start any such iteration, an initial estimate of the parameters is needed. One that exploits certain algebraic features of the model and is very simple is suggested in Luce (1959, pp. 32–34). This we used as our first estimate.

We decided that the final estimates should be ones that yield an approximate numerical minimization of χ^2. To calculate them, we used a computer program that starts with an initial estimate of the parameters and the observed choice frequencies as inputs, explores the 125 or 625 points of the parameter space that are generated by multiplying each of the initial parameter values by $1-2\varepsilon$, $1-\varepsilon$, 1, $1+\varepsilon$, and $1+2\varepsilon$ (where ε is a number between 0 and $\frac{1}{2}$), and then determines the coordinates of the point having the smallest value of χ^2. Depending upon the value of χ^2 from the initial estimates, we first used either $\varepsilon = 0.03$ or $\varepsilon = 0.075$. If necessary, further computer runs were conducted using the best current estimate of the parameters until the .03 grid had been used and the approximate minimum χ^2 parameters all lay interior to the boundaries of the grid. These estimates are reported below for subjects 1 and 2.

After these programs were prepared and applied to the data from several subjects, Dr. John van Laer (personal communication, 1961) worked out an analytic iterative scheme that could be used with a hand computer. Several comparisons were made between these estimates and the computer estimates, and it was found that they were very nearly the same. The estimates reported below for subjects 3–9 are based on van Laer's method.

3. Results

Before describing the main results of the study, it is necessary to eliminate from further consideration the data from subject 5 on the grounds of extreme instability, as judged, for example, by estimating his response proportions during successive fifths of the sessions. The variability of these estimates was excessive under the hypothesis of constant response probabilities. There is little or no trend in his data, indicating that the instability was not simply a case of learning. Since the other subjects were much more stable, the data for subject 5 are not discussed below.

Effect of irrelevant stimuli. Subjects 1 through 4 were run in conditions (1,2), (1,3), and (2,3), in which only a pairwise judgment was made, even though the three weights were hefted. Indeed, for subjects 1 and 2 these were the only two-response conditions run.[2] The estimated parameter values and the goodness of fit, omitting the M,M,M presentations, are shown in Table 3. Scatter diagrams of predicted vs. observed proportions for one of the better cases (Fig. 1) and for a poor one (Fig. 2) indicate the general quality of the prediction.

For 20 out of 24 cases in Table 3 (and in every case for the heaviest stimulus), the estimates of the stimulus parameters are smaller for the (i,j) conditions than for condition 2, suggesting that the irrelevant stimulus is having a definite effect not accounted for by the model. Because such an interaction was not of primary interest, the remaining subjects were not run on the (i,j) conditions.

[2] At the time the experiment was begun, we were trying to test directly the choice axiom given in Luce (1959, p. 6), which includes no provision for response-bias parameters; the response-bias model stated here had not then been developed. We conjectured that the choice axiom was likely to hold only when the stimulating conditions were constant. During our attempts to analyze these data—which clearly did not support the simple choice axiom—we developed the present model. This led us to run three more subjects, but we were still not confident that the new model would apply when the stimulating conditions as well as the response alternatives were varied, and so we continued to run the (i, j) conditions. It was only after we looked at these data in detail that we decided not to use these somewhat peculiar conditions with the remaining subjects.

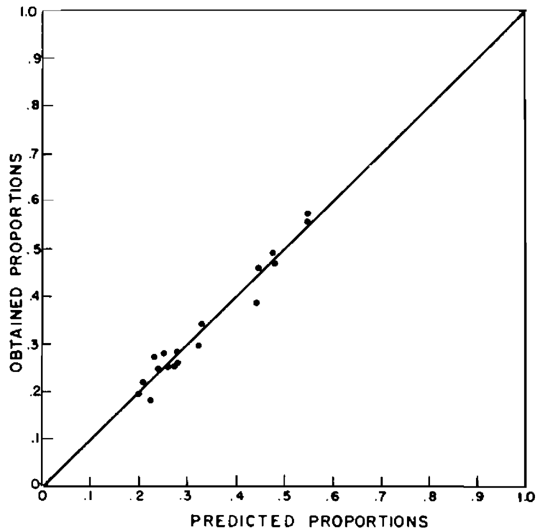

FIG. 1. Subject 3, Condition 2: obtained vs. predicted proportions, omitting *M,M,M* presentations.

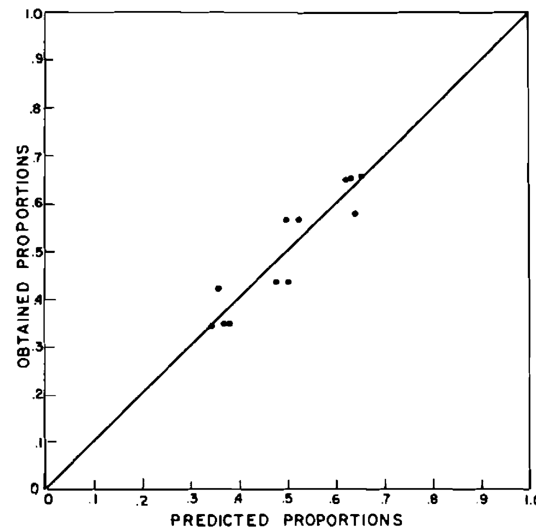

FIG. 2. Subject 3, Condition (1,3): obtained vs. predicted proportions, omitting *M,M* presentations.

TABLE 3

COMPARISON OF PARAMETER ESTIMATES AND GOODNESS OF FIT FOR CONDITIONS 2, (1,2), (1,3), AND (2,3), OMITTING *M,M,M* PRESENTATIONS

Subject	Condition	Parameters				Goodness of Fit		
		α	β	b_2	b_3	χ^2	df	p interval
1	2	3.63	1.26	1.27	1.45	19.92	8	.01, .02
	(1,2)	2.93	1.22	1.33	0	4.71	3	.10, .20
	(1,3)	3.11	1.31	0	1.20	22.09	3	< .001
	(2,3)	3.54	1.38	1	1.05	14.44	3	.001, .01
2	2	3.74	1.65	1.28	1.06	14.79	8	.05, .10
	(1,2)	3.13	1.62	1.16	0	2.07	3	.50, .70
	(1,3)	3.18	1.77	0	1.15	7.74	3	.05, .10
	(2,3)	3.15	1.54	1	.86	26.02	3	< .001
3	2	2.11	1.20	1.09	1.31	14.31	8	.05, .10
	(1,2)	1.75	1.08	1.03	0	15.29	3	.001, .01
	(1,3)	1.80	1.06	0	.95	18.77	3	< .001
	(2,3)	1.78	1.24	1	1.10	1.20	3	.70, .80
4	2	5.69	2.87	.84	.88	18.82	8	.01, .02
	(1,2)	5.12	2.45	.97	0	5.82	3	.10, .20
	(1,3)	3.97	2.33	0	.93	3.67	3	.20, .30
	(2,3)	5.19	2.64	1	.93	10.28	3	.01, .02

Comparison of primary conditions. As will become clear in the next subsection, the behavior of the subjects in the *M,M* and *M,M,M* presentations is quite different from their behavior in the other presentations, so we shall treat these cases separately. Thus, the comparisons we are about to make refer to conditions 1 and 2, omitting the *M,M* and *M,M,M* presenta-

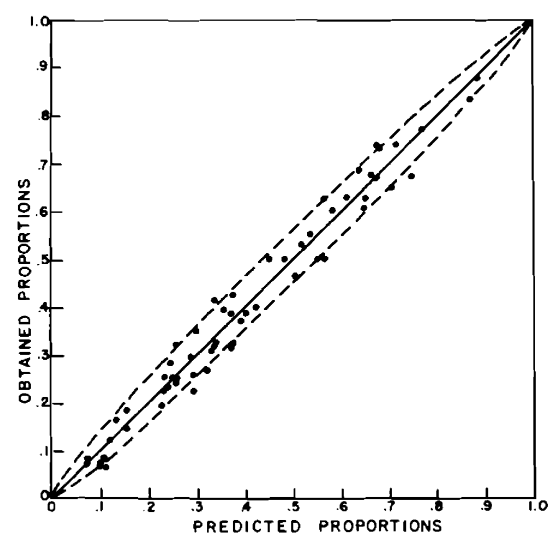

FIG. 3. Subject 6, Conditions 1, 2, 1*, 2*: obtained vs. predicted proportions, omitting *M,M* and *M,M,M* presentations. The same values of stimulus parameters and different values of response-bias parameters are used for each condition. The dotted curves are at ± two standard deviations for $N = 400$.

TABLE 4

COMPARISON OF STIMULUS PARAMETER ESTIMATES AND GOODNESS OF FIT FOR CONDITIONS 1, 2, 1*, AND 2*, OMITTING *M,M* AND *M,M,M* PRESENTATIONS

Subject	Condition	Parameters			Goodness of Fit		
		α	β	δ	χ^2	df	p interval
3	1	2.35	1.30	—[a]	9.03	3	.02, .05
	2	2.11	1.20	—[a]	14.31	8	.05, .10
	2*	—[a]	1.24	.741	2.89	8	.90, .95
4	1	5.32	2.60	—[a]	2.18	3	.50, .70
	2	5.69	2.87	—[a]	18.82	8	.01, .02
	2*	—[a]	2.55	.504	8.67	8	.30, .50
6	1	1.64	2.25	—[a]	8.73	3	.02, .05
	2	1.54	2.16	—[a]	18.75	8	.01, .02
	1*	—[a]	2.26	.382	17.93	3	< .001
	2*	—[a]	2.29	.275	34.84	8	< .001
7	1	2.45	1.55	—[a]	10.88	3	.01, .02
	2	2.53	1.75	—[a]	16.18	8	.02, .05
	1*	—[a]	1.49	.641	21.34	3	< .001
	2*	—[a]	1.53	.544	15.28	8	.05, .10
8	1	1.84	2.10	—[a]	4.69	3	.10, .20
	2	2.09	2.47	—[a]	11.85	8	.10, .20
	1*	—[a]	2.29	.690	16.03	3	.001, .01
	2*	—[a]	2.42	.706	27.24	8	< .001
9	1	2.82	1.78	—[a]	3.26	3	.30, .50
	2	2.81	1.74	—[a]	17.54	8	.02, .05
	1*	—[a]	1.49	.565	11.47	3	.001, .01
	2*	—[a]	1.57	.509	15.09	8	.05, .10

[a] Not obtained.

tions. In Table 4 the stimulus parameters and goodness of fit for subjects 3, 4, and 6–9 are shown for conditions 1, 2, 1*, and 2*. The response-bias parameters, which are discussed separately, are presented in Table 5.

To gain some sense of the adequacy of the model, we have prepared scatter diagrams of observed vs. predicted response proportions for the poorest subject, subject 6, and for the best, subject 9, using the geometric mean of the several estimates of the stimulus parameters and the separate estimates of the response-bias parameters. These are shown in Figs. 3 and 4.

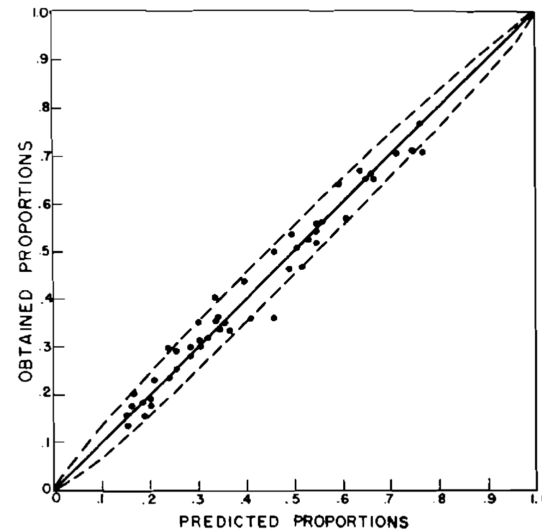

FIG. 4. Subject 9, Conditions 1, 2, 1*, 2*: obtained vs. predicted proportions omitting *M,M* and *M,M,M* presentations. The same values of stimulus parameters and different values of response-bias parameters are used for each condition. The dotted curves are at ± two standard deviations for $N = 400$.

Response-bias parameters. We can estimate the response-bias parameters in two ways, namely from the minimum χ^2 estimates based upon all presentations save *M,M* and *M,M,M*, and from these two presentations, which, according to the model, depend only upon the biases, not upon the stimulus parameters. These two sets of estimates are shown in Table 5.

Raw-response proportions. In order that these data can be used to test other models, the raw-response proportions are given in the Appendix.

4. Discussion and conclusions

The values of the stimulus parameters in Table 3 suggest that the stimulus parameters are not invariant when irrelevant stimuli are present. In light of experimental work on the effects of irrelevant stimuli on the precision of judgments (Gleitman, 1957) and theoretical work by Bush, Luce, and Rose (this volume, pp. 201–217) this is not surprising. Using a learning model for the bias parameters, the latter authors show that only under certain rather special experimental conditions—fortunately met in the remainder of this experiment—can we expect our simple model to describe discrimination data. We will not dwell further upon these data.

Turning to the data for conditions 1 and 2 and for their starred counterparts, we have two means of evaluation: goodness of fit and parameter invariance. If we approach the model as a null hypothesis, then in 14 out of 24 tests it must be rejected at the usual 5 per cent level of significance. Such

TABLE 5

Comparison of Response-Bias Estimates from Minimum χ^2 of All Presentations except M,M and M,M,M and from Those Presentations Alone

Subject	Estimate	Conditions 1	1*	2		2*	
		c_1	c_1	b_1	b_3	b_1	b_3
3	Min χ^2	1.12	—[a]	1.09	1.31	1.19	1.20
	M,M,M	1.04	—[a]	.957	1.34	1.39	1.61
4	Min χ^2	.86	—[a]	.840	.885	.769	.833
	M,M,M	.89	—[a]	.796	.730	.691	.926
6	Min χ^2	1.15	.94	1.40	1.53	1.33	1.31
	M,M,M	1.47	1.63	2.01	2.53	1.90	1.76
7	Min χ^2	.93	.96	.900	1.07	.918	1.04
	M,M,M	1.08	1.43	.946	1.22	1.74	2.13
8	Min χ^2	.95	.95	.786	1.08	.659	.885
	M,M,M	1.69	1.78	.983	1.69	1.16	1.89
9	Min χ^2	1.13	1.06	1.29	1.24	1.11	.953
	M,M,M	1.08	.99	1.66	1.37	1.28	.827

[a] Not obtained.

an approach to models is not, however, particularly useful, for all it means is that in a statistical sense the model fails to capture all that is present. The real question is how well what is present is captured, i.e., how well we can predict, not whether we can reject the hypothesis of perfect prediction. Figures 3 and 4 provide the best answer to this question: the model accounts for the data relatively well in the sense that there are no systematic deviations from the line of perfect prediction, and the clustering about that line is relatively tight (most points are within two standard deviations of the line). A comparison with other models for the same data would be instructive; however, we do not know of any that apply.

In Table 4, we have from one to four estimates of each stimulus parameter. Although there is some variability in these estimates, there does not seem to be any pattern to the irregularities. To get some idea about how much variability in these estimates is consistent with the hypothesis of invariance, consider the following example. Suppose that the true values are $\alpha = 2.5$ and $\beta = 1.5$ and that there are no biases. Then

$$p(1 \mid H,M,L) = \frac{\alpha}{\alpha + \beta + 1} = .50 .$$

A 10 per cent increase in the value of α changes this quantity from .50 to .524, and a 10 per cent increase in β changes it from .50 to .486. Since with 400 observations the standard deviation of estimates of $p = \frac{1}{2}$ is .025, both of these 10 per cent changes in the values of the parameters are within one standard deviation. The argument is clearly quite insensitive to the exact values of α and β within the observed range and to the exact biases, provided they are not very large. Therefore we conclude that parameter estimates that vary as much as 10 per cent are not the least bit surprising on the basis of the binomial variability of the estimates of the response probabilities. If, in Table 4, we calculate the mean estimate whenever there are two or more, we see that all estimates are within 10 per cent of the mean except for the two δ estimates for subject 6, which are 16 per cent from the mean, and one of the β estimates for subject 7, which is 11 per cent from the mean. Many of the estimates are much more tightly clustered than 10 per cent. We conclude, therefore, that the stimulus parameters exhibit the desired invariance over different experimental conditions.

It should be noted that $\alpha < \beta$ for subjects 6 and 8, an odd thing to happen since α was associated with a heavier weight than β. Apparently, in spite of the gloves, both subjects used additional cues (presumably small imperfections in the cylinder) along with their perceptions of weight in arriving at their responses. This was confirmed for subject 6 after the experiment was completed by substituting different containers. Because the main extra cue appears to have been associated with the M weight, which half the time was incorrect and half the time correct (conditions 1 and 2 vs. conditions 1* and 2*), their performance failed to improve as might otherwise have been expected. Were we interested in weight discrimination as such, this experimental error would invalidate these data and make the rest suspect; however, because our primary interest is in the model itself and because it simply does not matter when testing the model which cues the subject used or even whether they were correlated with the weights, the data are acceptable.

In contrast to the stimulus parameters, the response-bias parameters in Table 5 do not exhibit all the invariances that we should have liked. First, because the starred conditions were not identified to the subjects as being different from the unstarred ones, the bias parameters should be the same in both cases. The minimum χ^2 estimates are similar to about the same degree as the stimulus parameters; the only case where the departure exceeds 10 per cent of the mean value is for subject 9, bias b_3 in conditions 2 and 2*. In contrast, the M,M and M,M,M estimates are considerably more variable. Second, the M,M and M,M,M direct estimates of the bias parameters are quite different from those obtained from the rest of the data. In 26 of the 34 comparisons, the bias parameters from the all-M presentations are larger than the corresponding parameters estimated from the other six presentations (Minimum χ^2 in Table 5). This is the reason that we did not attempt to test goodness of fit with these presentations included.

The failure of the bias parameters to exhibit invariance casts considerable doubt upon the model. The difficulty seems to center primarily on the estimates obtained from the M,M and M,M,M presentations. These estimates are not consistent either when they are compared with the minimum χ^2 estimates in the same conditions or when the starred and unstarred conditions are compared. To be sure, the all-M estimates are based upon one-seventh of the data, and the minimum χ^2 estimates on six-sevenths. Nevertheless, estimates based upon 400 presentations should be better than this if the model is correct. Hence either the model is wrong or subjects were somehow able to distinguish the all-M presentations from the others. As we have seen, subjects 6 and 8 may have identified the M weight by some sort of tactile cues, and perhaps other subjects were also able to recognize the all-M presentations in some way and so behave differently.

For three responses, some evidence exists for the usual negative time-order effect in spite of the use of immediate feedback and rewards. When the minimum χ^2 bias estimates are examined, the third response tends to be more heavily used than the second, and the second more than the first (corresponding to $b_3 > b_2 > b_1 = 1$). However, when conditions with two responses are considered, half the subjects use the second response more frequently than the first in condition 1, but only one out of four does in condition 1*. Specific subjects seem to exhibit the same pattern in different conditions, e.g., note the preference of subjects 4, 7, and 8 for the first response. As noted above, the usual negative time-order effects are more pronounced for the bias parameters estimated from the all-M presentations than for those estimated from the other presentations by a minimum χ^2 technique.

5. Summary

Relative frequencies of choices of the heaviest of several lifted weights were determined for a variety of conditions using eight subjects. The purpose was to test a choice model of discrimination. The sets of weights or the possible responses, or both, varied from condition to condition. All possible orders of different weights were used within a condition. Approximate minimum χ^2 estimates of the parameters were found, and judging by scatter diagrams, the choice model accounts for much of the variability in the distribution of choices for different presentations within conditions. Different estimates of what should be the same stimulus parameter if the model is correct were obtained from different conditions. When no irrelevant stimuli were present, these different estimates were judged to be adequately similar in the sense that the deviations from the mean parameter value corresponded to something of the order of $1\frac{1}{2}$ standard deviations or less in the estimates of

the probabilities. When irrelevant stimuli were present, the estimates from the two-response conditions were considerably smaller than those obtained from the three-response situation, showing that the model does not apply when there are irrelevant stimuli. The minimum χ^2 estimates of the bias parameters were also judged to be adequately invariant across conditions; however, "pure" biases estimated from the responses to presentations of the same stimulus several times were neither consistent across conditions when they should have been nor consistent with the minimum χ^2 estimates. It is not known whether this failure can be attributed to experimental difficulties or whether it is sufficient to reject the model. Evidence of the usual negative time-order effects was found for some subjects.

Appendix

Relative frequency of choice of response

Note that one response is omitted in the following tabulations because the relative frequencies of all responses must sum to 1 for each presentation. Decimals are omitted.

1. Irrelevant stimulus conditions: entries are relative frequency of choice of ith response in condition (i, j).

Presentation	Condition (1,2)				Condition (1,3)				Condition (2,3)			
	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
H,M,L	627	600	568	694	713	750	660	788	598	706	525	775
H,L,M	686	748	620	899	613	561	580	648	468	478	440	328
M,H,L	208	306	335	381	576	620	565	744	765	782	630	847
L,H,M	211	216	342	208	453	363	565	330	672	654	550	643
M,L,H	502	571	562	766	218	306	348	375	176	284	332	177
L,M,H	407	355	528	357	186	211	347	205	250	289	388	318
M,M,M	—	—	530	603	—	—	498	539	—	—	512	530

2. Two weights presented: entries are relative frequency of response "One."

Pres.	Condition 1					
	S3	S4	S6	S7	S8	S9
H,M	581	682	371	609	446	569
H,L	680	865	601	702	672	702
M,H	305	356	501	370	521	341
L,H	262	173	394	291	363	235
M,L	568	759	674	674	703	640
L,M	455	331	260	450	360	351
M,M	490	529	405	481	372	480

Pres.	Condition 1*			
	S6	S7	S8	S9
M,L	651	552	667	567
M,VL	877	727	777	709
L,M	267	365	254	331
VL,M	165	293	266	296
L,VL	771	694	652	669
VL,L	327	430	441	353
M,M	380	412	360	502

3. Three weights presented: entries are relative frequencies of responses "One" and "Two."

Condition 2

Pres.	*S*1	*S*2	*S*3	*S*4	*S*6	*S*7	*S*8	*S*9
H,M,L	547,252	532,343	457,260	590,286	247,466	475,277	422,413	497,332
H,L,M	500,228	527,213	386,273	631,083	251,195	459,183	364,167	368,230
M,H,L	206,645	260,600	280,467	338,605	326,425	326,486	465,320	252,560
L,H,M	130,603	132,650	219,490	146,548	147,322	204,464	207,278	156,540
M,L,H	145,169	225,194	247,180	331,083	388,226	355,122	435,127	291,192
L,M,H	147,162	147,253	192,251	129,263	187,499	197,300	201,365	158,321
M,M,M	—	—	303,290	397,314	180,363	316,299	272,267	248,412

Condition 2*

Pres.	*S*3	*S*4	*S*6	*S*7	*S*8	*S*9
M,L,VL	381,348	667,219	500,413	489,318	620,190	505,361
M,VL,L	366,256	673,093	625,067	508,155	621,086	463,188
L,M,VL	299,441	296,599	195,736	304,518	341,502	314,553
VL,M,L	233,429	170,583	074,629	158,470	229,508	175,546
L,VL,M	303,285	266,094	254,077	335,141	255,131	315,154
VL,L,M	198,353	178,222	086,226	231,278	260,175	202,338
M,M,M	250,347	382,264	215,407	205,358	247,287	300,385

REFERENCES

Gleitman, H. Proactive and retroactive assimilation in the successive comparison of loudness. *Amer. J. Psychol.*, 1957, **70**, 117–119.

Green, D. M. Psychoacoustics and detection theory. *J. acoust. Soc. Amer.*, 1960, **32**, 1189–1203.

Licklider, J. C. R. Three auditory theories. In S. Koch (Ed.), *Psychology: a study of a science*. Vol. 1. New York: McGraw-Hill, 1959. Pp. 41–144.

Luce, R. D. *Individual choice behavior: a theoretical analysis*. New York: Wiley, 1959.

Luce, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963. Pp. 103–189.