METHODS & DESIGNS

Measurement of motivation and incentive

NORMAN H. ANDERSON

University of California, San Diego, La Jolla, California 92093

Many theories of motivation employ algebraic models in which motivation acts as an amplifier or multiplier. Such multiplying models can be used as the essential basis for measurement of motivation, incentive, and other theoretical constructs by means of a linear analysis from functional measurement. A numerical example is presented, together with discussion of various problems of data analysis. Opportunities, difficulties, and precautions are discussed for extending this approach to test the operant matching law, the postulates of Hull-Spence theory, and general problems of stimulus integration with behavioral response measures.

A continually attractive idea in motivation theory is that motivation acts as an amplifier or multiplier. This idea appears in Freud's concept of psychic energy, in Lewin's tension systems and his criticism that associationism lacks a "motor," and in Tolman's distinction between performance and learning. In the hands of Hull, this idea began to take on more quantitative form in the well known formula:

$$\begin{aligned} \text{Reaction Potential} &= \text{Drive} \times \text{Habit}, \\ E &= D \times H. \end{aligned} \quad (1)$$

Similar ideas, often in a more cognitive form, have been used by many other investigators (see, e.g., Klitzner & Anderson, 1977). For example, approach tendencies are generally assumed to follow the multiplying MEV model, Motivation $\times$ Expectancy $\times$ Incentive Value.

But the idea that motivation acts as a multiplier has remained quasimathematical, not a true equation. The reason is simple: No adequate way to measure motivation and the other terms has been developed. Without a capability for measuring psychological values on linear ("equal-interval") or ratio scales, such equations remain hopeful verbalisms.

Hull was keenly aware of this problem, for his entire theory revolved around algebraic equations for behavior. In his idea book for May 24, 1945, Hull (1962) writes that he seems at last to have devised a technique that will "probably be able to penetrate the hitherto

inaccessible ring of the constants of my system in such a way as to measure s_{HR} , D , s_{ER} , ..." (p. 871). According to Hovland (1952), Hull considered his experiments on quantification methodology as one of his most significant contributions. However, this attempt to adapt Thurstonian methods to scale runway speed in terms of its variability does not seem to have been successful.

Anderson (1962) proposed an alternative approach that bypassed Thurstone's variability units to place measurement directly within the structure of the theoretical laws. "The basic idea of the present method is to use the theoretical equations directly in the scaling procedure" (p. 408).

Stimulus Integrations

A central theme in the functional measurement approach is that of stimulus integration. Most if not all behavior represents the combined or integrated resultant of several coacting stimuli. Functional measurement methodology has developed as part of a research program on general problems of stimulus integration (Anderson, 1974a, p. 215, 1974c, p. 236).

Two main operations arise in stimulus integration, namely, integration and valuation. Integration refers to the combination rules that govern the integration of separate discrete stimuli into a unitary response. Valuation refers to the processes that transform the physical stimuli into their subjective psychological counterparts.

The close relation between valuation and integration becomes explicit in any algebraic integration model. The algebraic structure of the model represents the integration operations. The terms of the model represent the outcome of the valuation operations. The concept of stimulus value is thus inherent in the statement of the model, and the model analysis is closely connected to problems of stimulus scaling.

The traditional approach has viewed scaling as a

I wish to express my thanks to Michael Klitzner for the opportunity to use some of his thesis data, to Edmund Fantino and Frank Logan for helpful comments, and to Joseph Farley for a valuable critique of an earlier draft. This work was supported by National Science Foundation Grants BMS 74-19124 and BNS 75-21235. Support was also provided by grants from the National Institute of Mental Health to the Center for Human Information Processing, University of California, San Diego. Reprint requests may be sent to Norman H. Anderson, Psychology C-009, University of California, La Jolla, California 92093.

preliminary to the study of substantive problems. This approach is well illustrated in applications of Thurstonian scaling (as in Hull's case cited above) that attempt to obtain stimulus scales from variability measures that are separate from the behavior under study. This direction of attack is obvious and natural, seemingly similar to that employed in physical science. Progress would certainly be much facilitated if valid stimulus scales were available at the beginning of an experiment. Unfortunately, this direction of attack has not worked very well in psychology.

Functional measurement reverses the traditional approach to center directly on the study of the integration model. This model becomes the base and frame for scaling. The integration-theoretical approach has had considerable success across a diverse array of substantive areas (Anderson, 1974a, 1974b, 1974c). The present article shows how this same approach might be applied to the study of motivation.

A Hypothetical Experiment

A hypothetical experiment on effects of motivation and valence is shown in Table 1. The three rows represent three levels of thirst. The four columns represent four thirst quenchers: warm water, cold water, Coke, and beer. Thus, there are 12 experimental conditions in this two-factor design. The numerical entries represent how much a person is willing to pay under each condition.

The question is: Do these data obey the multiplying model,

$$R = C_0 + M \times V \quad (2)$$

Here, R is the observed response measure, M denotes the psychological value of the motivational factor, and V denotes the psychological value of the incentive factor. The term C_0 is included to allow for an arbitrary zero in the response scale.

This question is almost trivial in principle. Equation 2 says that the response is a straight-line function of V , with slope equal to the value of M . All that is needed, therefore, is to plot each row of data in Table 1 as a function of the V values of the four thirst quenchers. Each row should plot as a straight line, and so the

Figure 1. Factorial plot of hypothetical data of Table 1. Linear fan pattern reveals the multiplicative form of the integration function.

several rows should form a linear fan. This linear fan shape is the sign of a multiplying model, so this plot provides a simple graphical test of the model.

But one vital element is lacking: The V values of the four thirst quenchers will not ordinarily be known. Without them, little can be done and the situation might seem hopeless.

But, in fact, the V values are implicit in the data. If the model does apply, then the column means of Table 1 constitute a linear scale of the V values (see below). Accordingly, these column means may be used as provisional estimates of the V values.

That has been done in Figure 1. The four liquids have been spaced on the horizontal axis at the values prescribed by the column means of Table 1. Each row of data does plot as a straight line. It would be concluded, therefore, that these data do obey the motivation $\times$ valence model.

Moreover, the success of the model validates the provisional scaling of valence. The psychological values of motivation may be obtained similarly from the row means of the data table. This functional measurement analysis thus solves two difficult problems in a simple way. Not only has the model been verified, but the analysis provides validated linear scales of the two stimulus variables, motivation and valence.

An Experiment on Anticipated Fear

In this experiment, a preliminary study in an investigation of snake phobia (Klitzner, 1977), subjects used a graphic rating scale to indicate how frightening they would consider such situations as "There is a rattler in your backyard." One of the designs that was used is shown in Table 2. Each of four species of snakes is combined factorially with each of four physical locations. The entries in the table are the mean ratings, averaged over two independently randomized

Table 1
Amount Paid for a Drink as a Function
of Motivation and Incentive

Thirstiness	Thirst Quencher			
	Water		Coke	Beer
	Warm	Cold		
Slight	13	15	21	17
Moderate	15	21	39	27
Great	20	36	84	52
Mean	16	24	48	32

Table 2
Judged Fearfulness of Snake-Proximity Situations

	Snake			
	Garter	Boa	Rattler	Cobra
Hold in hand	2.9	11.0	12.1	14.3
In backyard	2.5	8.6	9.6	11.9
In neighborhood	2.5	7.5	8.2	9.3
In the world	1.4	3.7	3.7	4.4
Mean	2.3	7.7	8.4	10.0

presentations of the 16 stimulus combinations for each of 10 unselected undergraduates.

The theoretical hypothesis is that the two stimulus variables combine by a multiplying rule:

$$\text{Fear} = \text{Proximity} \times \text{Snake.} \quad (3)$$

Here, snake corresponds to the valence term of Equation 2, while proximity corresponds to an expectancy term.

The analysis of these data proceeds in the manner outlined above. First the column means are calculated; these are the provisional fear values of the four snakes. The four snakes are placed on the horizontal axis in Figure 2, spaced out according to the provisional fear values. Then each row of data from Table 2 is plotted as a separate curve.

The theoretical linear fan shape is quite clear in Figure 2. Some discrepancies can be seen, but these are no more than expected from prevailing response variability.

A different group of 10 subjects was run in a fairly similar expectancy $\times$ valence design. These data, treated in the same way, also exhibit the linear fan form, as shown in Figure 3. The fear values of the snakes are again measured by their spacing on the horizontal axis.

Figure 2. Factorial plot of anticipated fear as a function of species and nearness of snake. Linear fan pattern reveals the multiplicative form of the integration function.

Figure 3. Factorial plot of judged danger as a function of probability of given snake being in the room. Linear fan pattern reveals the multiplicative form of the integration function.

These values are somewhat different in the two figures, in part because of individual differences, in part because of within-individuals response variability.

As far as one experiment goes, therefore, these data have accomplished two goals: They support the multiplying model. In addition, they provide validated linear scales of the subjective values of the two stimulus variables.

The simplicity and directness of this analysis are notable. No prior stimulus scaling is required. For the present data, it seems enough to plot the raw data and look at the graph. In general, of course, a statistical test of goodness of fit is required, as discussed and illustrated elsewhere (e.g., Klitzner & Anderson, 1977).

THE LOGIC OF FUNCTIONAL MEASUREMENT

The essential logic of functional measurement "consists in using the postulated behavior laws to induce a scaling on the dependent variable" (Anderson, 1962, p. 410). The term "functional" derives from the behavior law considered as a mathematical function. Also, the term emphasizes that the resultant stimulus values are those that were functional in the observed behavior.

Of course, success depends completely on the empirical validity of the model. It is the model or function that provides the base and frame for measurement. Accordingly, the model analysis requires more detailed consideration, especially with respect to two implicit assumptions that were passed over in the above examples. This can most conveniently be done by beginning with an explicit statement of the linear fan theorem.

Linear Fan Theorem

As shown in the above numerical examples, the

analysis of the multiplying model rests on the linear fan shape demonstrated in Figures 1-3. This is made explicit in the following linear fan theorem:

Suppose that (1) the multiplying model of Equation 2 is true, and that (2) the observed response measure is a linear scale. Then (1) the appropriate plot of the raw response data will form a set of diverging straight lines, and (2) the marginal means of the factorial design will be linear scales of the psychological values of the stimuli.

Both assumptions are necessary. If one assumption was not satisfied, then the linear fan property would not in general be obtained. If both assumptions hold, then the psychological stimulus values are obtained from the marginal means, as illustrated in Table 1 and proved below. The linear fan property may then be derived by the reasoning in the above examples.

It follows, of course, than an observed linear fan provides joint support for both assumptions of the theorem. Thereby, it accomplishes three simultaneous goals: (1) It supports the multiplying model; (2) it supports the linearity of the response measure; and (3) it provides validated linear scales of stimulus value.

It deserves emphasis that two distinct measurement problems are involved, of the response and of the stimuli. Both depend on the stimulus integration function, in this case the multiplying model.

Linear Response Measures

The success of the rating response in the above experiment is consistent with previous work on information integration theory. Numerous experiments across many different experimental tasks have shown that only modest experimental precautions (Anderson, 1974a, p. 231, 1974c, p. 245) are needed to linearize the rating response.

In general, of course, overt response measures cannot be expected to be linear scales of covert response strength. More behavioral response measures, such as nearness of approach to the snake, or some physiological measure might well be no more than ordinal scales, nonlinearly related to the true measure of fear. The same holds for common response measures in animal research, such as runway speed or barpress rate. In this case, the above analysis will not work. Fortunately, this case also has a straightforward solution, as shown below.

Stimulus Independence

Equation 2 contains the implicit assumption that the two stimulus variables have independent effects. In terms of factorial design, this assumption means that the psychological value of a given row stimulus is constant, independent of the column stimulus, and vice versa. This independence assumption is standard in the study of mathematical models. Usually, the independence assumption is considered as part of the model, though it should, in principle, be separated from the assumption about the mathematical form of the integration function.

Motivational factors might perhaps be expected to affect the valence of the incentive object. For example, beer will taste better to the more thirsty person, and that might seem to reflect a change in stimulus value. However, the model does allow for such an effect, since the motivational state acts as a multiplier on the valence of the incentive. There will be a statistical interaction because the curves are nonparallel. However, the linear fan shape supports the independence assumption that the stimulus parameters are constants. The integration operation might be considered to be interactive, but not the stimulus valuation.

Of course, little is known about stimulus interaction in the area of motivation, and such effects may well be frequent. However, the work on algebraic models in other areas has given broad support to the independence assumption (e.g., Anderson, 1974a, 1974c). Even if independence holds only under restricted conditions, it can still be very useful in experimental analysis.

Two Qualifications

Both assumptions of the linear fan theorem are necessary. If just one was wrong, then the linear fan pattern would in general be destroyed. It is a logical possibility, however, that both assumptions are wrong but just cancel each other to leave a linear fan. This possibility does not seem especially likely in general, although it could escape detection in a small experiment. However, an interesting case does arise in connection with monotone response scaling, as discussed below.

A related qualification is that any one experiment only goes so far. Functional measurement occurs as part of some substantive theoretical system. Confidence in the system requires an interlocking body of concepts and data from many experiments. This problem of generality is discussed elsewhere (Anderson, 1977b, p. 212).

SOME TECHNICAL CONSIDERATIONS

A few pertinent technical aspects of functional measurement methodology will be discussed briefly here. More exhaustive treatment can be found in the references cited above.

Derivation of the Stimulus Scales

The linear fan theorem states that the column means provide a linear scale of subjective valence. This statement is easy to prove, and the proof brings out one or two ancillary points of interest. For simplicity, response variability is neglected.

Let M_1, M_2 , and M_3 denote the values of motivation for the three rows of Table 1, and let $\bar{M} = (M_1 + M_2 + M_3)/3$ be their mean value. Similarly, let V_1, V_2, V_3 , and V_4 denote the values of valence for the four columns. Then the mean for the first column is $1/3[(C_0 + M_1 V_1) + (C_0 + M_2 V_1) + (C_0 + M_3 V_1)] =$

$1/3(C_0 + C_0 + C_0) + 1/3(M_1 + M_2 + M_3)V_1 = C_0 + \bar{M}V_1$. A similar expression holds for each other column. Thus, the four column means can be written as $C_0 + \bar{M}V_1, C_0 + \bar{M}V_2, C_0 + \bar{M}V_3, C_0 + \bar{M}V_4$.

These four column means are a linear function of the V values. In other words, the column means constitute a linear ("equal-interval") scale of the true subjective valences. Similarly, the row means constitute a linear scale of the subjective values of motivation. Clearly, this method of proof applies to any factorial design.

The terms C_0 and $\bar{M}$ correspond to the zero and unit of the valence scale. The zero and unit are arbitrary, of course, just as in the Fahrenheit and Celsius scales of temperature. The unit, in particular, is a function of the motivation values that happen to prevail in any case.

Statistical Test

In some situations, the graphical test of the linear fan property outlined above will be sufficient. Ordinarily, however, an adequate test of the model requires a more formal statistical assessment that takes account of response variability. That is reasonably straightforward, using analysis of variance.

The model test rests on the interaction term in the analysis of variance. If the multiplying model is correct, then the systematic part of the interaction will be concentrated in a single degree of freedom, corresponding to its linear $\times$ linear component. The residual interaction should be nonsignificant, therefore, being composed merely of statistical variability.

In Figure 2, for example, 96.4% of the interaction sum of squares was concentrated in the bilinear component. Only 3.6% remained for the whole residual, or less than .5% per degree of freedom. In Figure 3, 85.8% of the interaction sum of squares was concentrated in the bilinear component, leaving 2.8% on a pro rata basis for each residual degree of freedom.

These variance percentages might almost seem to eliminate the need for formal statistical tests. Unfortunately, variance percentages can be extremely misleading (Anderson, 1977b; Anderson & Shanteau, 1977; Footnote 1 below). Indeed, visual inspection of the factorial plot may well be superior. Exact statistical tests are available and in general they are necessary. It should be noted that repeated-measurements design produces a bias in the test of the residual (Anderson, 1976, Footnote 1, 1977b, Footnote 5) that may be avoidable with individual-subjects design (e.g., Klitzner & Anderson, 1977).

Individual-Organism Analysis

Functional measurement is applicable to the individual organism. This is important because individual differences in motivation and incentive value tend to be great, especially for aversive stimuli. An adequate theory must be able to work in terms of the value system of the individual organism.

In practice, of course, that ordinarily requires that the individual be tested in all cells of the factorial design (or some reasonable facsimile thereof). Such repeated testing will not always be experimentally feasible. If the data are available, however, then the above methods can be used for each individual. The functional scaling will then yield the personal values of that individual.

Question of Demand Effects

Judgmental studies with rating responses, as in Figures 2 and 3 above, sometimes cause an uneasy concern that the predicted linear fan is somehow imposed by the demands of the experimental task, facilitated especially by the practice of running each subject in all stimulus combinations.

It is true, of course, that there is an intention and hope that the subject will use the overt response measure as a linear scale of subjective feeling. Experimental procedures are chosen with this end in view. The important point, however, is that these procedures do not impose any demand on how the stimuli are combined or integrated. In terms of the functional measurement diagram (Anderson, 1974a, p. 281, 1977b, p. 202), the constraint is on the response output function, not on the stimulus integration function.

One line of evidence against the demand interpretation comes from studies of the psychophysical law. Studies of heaviness have obtained the same scale from two quite different tasks, weight averaging and size-weight illusion (Anderson, 1974a, p. 235). The size-weight task does not suffer from a demand problem, since the subjects are typically unaware that the size cue is affecting their judgments. Accordingly, the cross-task agreement of the sensory scales is evidence against a demand interpretation of the weight-averaging task. The same conclusion is supported by the cross-task agreement of grayness scales obtained from bisection, averaging, and differencing tasks (Anderson, 1977b, Figure 7). Many other lines of evidence support the same conclusion.

GENERAL ALGEBRAIC MODELS

Algebraic models of behavior have been considered by numerous writers, and some basic forms other than the multiplying model need brief mention.

Linear Models

Linear models include those in which the stimulus variables combine by adding or subtracting operations. For example, the incentive value of a meal might be the sum of the incentive values of the component foods (Klitzner & Anderson, 1977; Shanteau & Anderson, 1969). Similarly, the net response tendency in a conflict situation might be conceptualized as the difference between approach and avoidance tendencies (Anderson, 1962).

Functional measurement methods provide a very simple analysis of the linear model. Graphically, linear models are characterized by a property of parallelism, analogous to the linear fan property of multiplying models. Indeed, the above linear fan theorem can be changed to a corresponding parallelism theorem by changing just two phrases, from multiplying model to linear model in the first premise, and from "diverging straight lines" to "parallel curves" in the first consequence.

This graphical test of parallelism is even simpler than the linear fan test. The statistical analysis is also simpler. The linear model implies that the interaction term in the analysis of variance is zero in principle and nonsignificant in practice. This provides a powerful tool for studying incentive integration.

Averaging Model

Averaging processes seem to be ubiquitous in stimulus integration (e.g., Anderson, 1974c). Many situations that have seemed at first to call for an adding model have turned out to obey an averaging model.

This point is important for two reasons. Methodologically, the averaging model is linear only under the restriction that all the row stimuli have equal importance or weight among themselves, and similarly for the column stimuli. When this equal-weight condition does not hold, systematic deviations from parallelism are expected and the model analysis becomes substantially more difficult. To take advantage of the parallelism property, therefore, it can be worth considerable effort to try to insure the equal-weighting condition.

Theoretically, the averaging model has the remarkable property that it can provide direct comparison between stimulus variables that are qualitatively different (Anderson, 1974a, p. 227; Norman, 1976). Under appropriate conditions, scale values of different stimulus variables will be estimated on linear scales with common zero and common unit, and weights will be estimated on ratio scales with common unit. This property opens up the possibility of valid quantitative comparisons between qualitatively different drive states, such as hunger, fear, and sex, and between qualitatively different incentives, such as food, drink, and money. It should be emphasized, however, that special design and analysis are necessary.

Ratio Model

Experiments in several areas (e.g., Anderson & Farkas, 1975; Leon & Anderson, 1974; Oden, 1974, 1977) have supported a ratio model of the form,

$$R = X/(X + Y). \quad (4)$$

The common characteristic of these diverse experimental studies appears to be that the response depends on compromise between two competing response

tendencies, here denoted by X and Y. Accordingly, this ratio model may arise frequently in behavioral studies of learning and motivation.

Theoretically, the ratio model can be derived from averaging theory, as noted below. It has the same form as Luce's (1959) choice model. However, Luce's model deals with probabilistic choice responses, whereas the present ratio model allows for deterministic responses on a continuous scale.

Diagnostics for Compound Models

When more than two stimulus variables are involved, both adding and multiplying operations could play a role. Functional measurement provides a diagnostic tool to uncover which, if any, multilinear model underlies the response. Two (or more) variables that add should exhibit the parallelism property; two or more variables that multiply should exhibit the linear fan property. These two properties allow one to decide among various possible models.

This diagnostic capability is illustrated in Table 3. With three stimulus variables, there are five possible multilinear models, excluding linearly equivalent forms. These are listed in Table 3, together with their respective interaction structures. Each model has a unique pattern of interaction. Hence, the observed pattern of interaction can be used to determine which, if any, multilinear model underlies the response (Anderson, 1974a, p. 264).

In Model 3, $X(Y + Z)$, for example, two interactions, namely, X by Y and X by Z, should be nonzero and should exhibit the linear fan property. The other two interactions, namely, Y by Z and X by Y by Z, should be zero and so should exhibit parallelism. If these properties do hold in fact, then Model 3 must be the operative model. An application to Hull-Spence theory of motivation is given below.

This example shows how the observed interaction structure can be used for diagnostic purposes. The same approach can be used when more than three stimulus variables are involved. Pertinent experimental applications are given by Klitzner and Anderson (1977) and by Shanteau and Anderson (1972).

Krantz and Tversky (1971) have attempted to apply conjoint measurement for diagnosis of multilinear models. A basic assumption of conjoint measurement is that numerical response measures are not allowed.

Table 3
Diagnostics for Three-Variable Multilinear Models

Model	Interactions		
	Bilinear	Trilinear	Zero
1. $X + Y + Z$			XY, XZ, YZ, XYZ
2. $XY + Z$	XY		XZ, YZ, XYZ
3. $X(Y + Z)$	XY, XZ		YZ, XYZ
4. XYZ	XY, XZ, YZ	XYZ	
5. $XY + XZ + YZ$	XY, XZ, YZ		XYZ

Only ordinal or directional relations among the response data may be used in model analysis. The advantage of such tests is that they require only an ordinal or monotone response scale. The disadvantage is that they are not very practicable for data analysis.

The problem is that conjoint measurement assumes zero response variability and lacks an error theory. Ordinal or nonmetric analysis is generally practicable only when each stimulus variable has a number of levels that are fairly closely spaced. But then conjoint measurement properties such as double cancellation will have numerous violations due merely to response variability. Without an error theory, it is generally difficult to guess whether any observed frequency of violations is systematic or merely chance. As a consequence, applications of conjoint measurement have been extremely rare (see, e.g., Cliff, 1973, p. 478; Falmagne, 1976, p. 65; Tukey, 1969, p. 88).

Within the functional measurement approach, however, a valid error theory for ordinal analysis has been developed. Accordingly, the above diagnostic tools remain usable even when monotone transformation of the response is allowed.

Compound Models and Response Scaling

Multioperation models can be useful when the linearity of the response scale comes under question. Model 2, $XY + Z$, contains both multiplying and adding operations. Suppose that the linear fan test infirms the multiplying operation, but that the parallelism test supports the adding operation. Of itself, the failure of the multiplying operation might be attributable to failure of the model, or to response nonlinearity. However, success of the adding operation tends to rule out response nonlinearity.

A related application would be to include an already established integration operation, not perhaps of interest in itself, as a basis for response rescaling. This established operation could serve as a frame for transforming a possibly monotone response into a linear scale. The interpretation of the data with respect to the other integration operation need not then be troubled by uncertainty about the response scale. The potentialities of this application indicate the premium on establishing some integration operation, no matter how prosaic, in any area of study.

Experimental applications of these and related ideas are given by Anderson and Cuneo (in press), Birnbaum and Veit (1974), Klitzner (1975), and Shanteau (1974a, pp. 251ff, 1974b, pp. 21ff, 1977b, pp. 210-213). Applications to Hull-Spence theory and the problem of monotone indeterminacy are considered below.

NONLINEAR RESPONSE AND MONOTONE TRANSFORMATION

When the observed response measure is a linear scale, the analysis of the algebraic models is simple

and direct. But, in general, the observed response cannot be expected to be a linear scale. An adequate theory of measurement must allow for response nonlinearity.

Monotone Rescaling Principle

The problem of response nonlinearity arose in the initial theoretical article on functional measurement (Anderson, 1962). That article showed how factorial design led to a simple parallelism test for an approach-avoidance model without any need for prior scaling of the stimulus variables. However, the parallelism test does assume a linear response scale, an assumption that had uncertain validity for response speed.

In principle, however, the problem of nonlinear response has a straightforward solution. Under reasonably general conditions, the observed response should be at least a monotone ("ordinal") scale, monotonically related to the true response measure. Hence, there is some monotone transformation that will linearize the response, transforming it from a monotone to a linear ("equal-interval") scale. If the model is correct, then this monotone transformation can be computed by using the constraints imposed by the experimental design. For the additive model, in particular, the desired monotone transformation is that which induces parallelism in the factorial plot. If the model is not correct, then the best possible monotone transformation would not generally be able to produce parallelism (see also below).

Three Technical Problems in Monotone Response Transformation

The first practical problem in monotone response transformation is to compute it. The original approach based on power series (Anderson, 1962) was developed mathematically by Bogartz and Wackwitz (1971) and computerized as Weiss' (1973) FUNPOT program. However, this approach has not been as useful in practice as the nonmetric monotone regression analyses embodied in Kruskal's (1965) MONANOVA and the more recent ADDALS of de Leeuw, Young, and Takane (1976). These two nonmetric programs find the monotone transformation that makes the data fit an additive model as well as possible.

At this point, the second problem arises; namely, testing whether the best possible monotone transformation is good enough. If not, then systematic deviations from parallelism should remain in the transformed data. Unfortunately, no test for this has previously been available because the transformation uses up an unknown number of degrees of freedom (Weiss & Anderson, 1972).

Fortunately, a recent development in functional measurement methodology has shown how this problem of testing goodness of fit can be solved in a rather simple way. Valid tests of additive models can thus be obtained even with monotonically transformed response data. The logic and some experimental applications are discussed elsewhere (Anderson, 1977a, 1977b).

The third problem remains in an unsatisfactory state. This problem is that the monotone transformation has great flexibility and so may be too obliging by fitting the data to the model even though it is wrong. That can happen when the experimental design does not provide sufficient constraint. In a 2 by 2 design, for example, the response can always be monotonically transformed to additivity unless there is an actual crossover interaction. Unless the experimental design provides sufficient constraints, therefore, a successful monotone transformation may mean very little. On this problem, unfortunately, only scraps of information are available (Anderson, 1977b) and it requires detailed study.

Monotone Indeterminacy

One final problem with monotone transformation requires brief mention. To illustrate, suppose that $C_0 = 0$ in Equation 2 and that no values are negative. Then a log transformation on the response transforms the given multiplying model to an adding model. These two models are monotonically equivalent, therefore, and cannot be distinguished without some criterion external to the given data.

One way to avoid monotone indeterminacy is by use of compound models. Thus, the log-transform trick does not work with the compound model $(XY + Z)$. A relevant experimental example is given by Klitzner and Anderson (1977) and various other considerations are discussed elsewhere (Anderson, 1974a, p. 230, 1977b, pp. 210-213).

ALGEBRA OF JUDGMENT

Cognitive Algebra

Research in judgment theory has found extensive evidence for the operation of simple algebraic models. Successful applications have been made in many areas, including psychophysics, decision theory, and social attitudes (Anderson, 1974a, 1974b, 1974c). Not all judgment tasks obey a simple model, of course, nor should they be expected to. However, the recurrence of algebraic integration functions in all these areas suggests the operation of a general cognitive algebra.

A not infrequent objection to judgmental studies is that what a person says often has little relation to what he does. This has given rise to some prejudice against verbal judgments, in favor of "actual behavior."

From a cognitive view, however, motivational judgments are important. People are continually forming intentions and making promises, to themselves as well as to others. The operation of the human world, from installment buying to marriage, depends heavily on such intentions and promises.

Intentions and promises are, of course, only part of the forces that operate in any situation. A person may promise himself to touch the snake and pick it up next time, only to find his intentions overwhelmed when he again nears the snake. Nevertheless, his present

judgments of his intentions and motivations reflect one determinant of his future action. Such verbal behavior is relevant to nonverbal behavior and, of course, has its own importance.

Judgmental Behavior

In humans, there is no sharp line dividing "judgment" and "behavior." Indeed, it seems safe to say that judgmental processes underlie most behavior. Illustrative studies of judgmental behavior are provided in the dissertations of Klitzner (1977) and Lopes (1976).

In Klitzner's (1977) study of snake phobia, subjects received information about the aversiveness of the snake stimulus on each trial, and about the likelihood that they would contact the snake. To avoid contact, the subjects could purchase marbles from a gumball-type machine, in which a special gold marble would prevent any contact with the snake stimulus. The amount of money spent in attempted avoidance followed the postulated expectancy $\times$ valence model.

Moreover, individual differences in snake fear acted as a multiplicative motivational factor. This result illustrates a new approach to the study of individual differences as a motivational variable.

In Lopes' (1976) study of poker, long-term subjects played a modified, computerized version of five-card stud. Every individual subject obeyed a multiplying, expectancy $\times$ expectancy model suggested by probability theory. The linear fan was observed not merely for rated likelihood of winning but also for amount bet. The latter behavior was notable because it was a money-losing strategy. Since the bets were for even money, the subject should rationally have bet nothing when he felt he had less than a 50:50 chance to win. Even for these fairly skilled subjects, however, amount bet appeared to be directly proportional to expectancy of winning. Behavior and judgment both followed the same law.

ALGEBRA OF BEHAVIOR

Many writers have postulated that nonverbal behavior will obey simple algebraic models. Hull is the most prominent, but many other theorists have operated in the same vein. However, analysis of these models has been held back because of a lack of measurement capability. This problem, illustrated in the quotation from Hull (1962) in the Introduction, has remained largely unresolved.

Unfortunately, experimental analysis is substantially more difficult for algebraic models of behavior than for the judgmental models considered above. These difficulties arise from the nature of the experimental tasks that are typical of behavioral studies, including the concentration on animal subjects. These methodological problems deserve preliminary consideration.

Methodological Considerations

Need for linear response measures. A basic problem

for behavior theory revolves around the development of linear scales of response. Suppose that some behavior is expected to obey an additive model, say, but that the data show deviations from additivity. If these deviations merely reflected nonlinearities in the response scale, then they could be eliminated by monotone transformation. However, the deviations could also reflect an incorrect model, or some failure of the assumption of stimulus independence, in which case monotone transformation would be inappropriate. Further, some mixture of all three faults could be present, a situation that would be especially difficult to sort out.

The rapid progress on problems of stimulus integration in judgment theory has only been possible because the ordinary rating scale can, with modest experimental care, perform as a linear response scale. This greatly facilitates the analysis of models as well as general problems of stimulus integration.

In behavioral studies, it is even more important to develop linear response measures. They can be invaluable for study of the integration models, and even more for the study of stimulus interaction. Linear behavioral scales may be obtainable only in certain tasks, and only under certain experimental conditions. Even if they can be established only under restricted conditions, their value as an opening wedge can hardly be overestimated.

Design considerations. The designs that arise in behavioral and judgmental studies show two important modal differences. First, behavioral studies are frequently on a between-subjects basis, whereas judgmental studies tend more toward within-subjects design. High variability from individual differences reduces the power of between-subjects design to detect deviations from the model. That favors erroneous acceptance of the null hypothesis that the model is correct (Anderson, 1977b, p. 208). For this reason, between-subjects design has very limited value in model analysis, at least when the linear nature of the response measure is not yet established.

The second main point of difference relates to design size. In judgmental studies, a 5 by 5 design is typically easy to use and not especially large. In behavioral studies, a 5 by 5 design would be exceptionally large. If the response measure is a linear scale, then small designs may be adequate. However, if the response scale is only monotone or ordinal, then a 5 by 5 design may be too small to justify application of linearizing transformations.

Both of these factors are important in the selection of experimental tasks for studying algebraic models of behavior. Desirable task characteristics include rapid adjustment as subjects are transferred across conditions, reversible steady-state behavior, and flexibility in choice of stimulus variables. These and other considerations will appear in the subsequent discussion.

Behavioral measures. There is some general feeling that certain behavioral measures, such as runway speed

or press-peck rates, may be linear or near-linear response scales, at least under certain conditions. However, there are basic differences between such behavioral measures and the verbal ratings typical of judgment theory.

Ratings can be considered as pointer readings. A judgment of loudness, for example, is clearly distinct from the loudness itself. Although the response process itself is not without interest, primary concern is not with the response but what it points to, namely, the loudness. Moreover, these pointer readings have similar quality across the scale range.

Standard behavioral measures are different in two respects. First, they are magnitude responses, requiring different degrees of effort for different rates. High speed and rate are physiologically different from low speed and rate. For speed measures, especially, some kind of law of diminishing returns seems not implausible. The observed measure may thus be only a monotone scale. This issue arises below in discussion of the additive results of Logan and of Guttman.

A more serious problem is that the response measure may be nonhomogeneous, a molar aggregate of distinct response classes. Logan (1960, Chapter 6) gives a sophisticated discussion of the problem of response definition for runway studies with special reference to response-correlated reinforcement and molar-molecular distinctions. Shimp (1975) gives a wide-ranging discussion of operant studies with particular concern for molecular analysis. The importance of this problem is well illustrated by human probability learning in which the shift from mean response to sequential dependencies was theoretically decisive (Anderson, 1959, 1964, 1974c, p. 267; Jones, 1971; Myers, 1976). This problem cannot be pursued here, but it may be noted that model analysis can provide one basis for attaining construct validity. If the response measure is quantitatively adequate, that supports its qualitative integrity.

Three Prototype Studies

Guttman. The report by Guttman (1954) illustrates many of the features that seem desirable in animal studies of integration models. Subjects were hungry rats who pressed a bar to obtain one of seven concentrations of one of two sugar solutions (sucrose or glucose) on a variable-interval (VI) schedule. Guttman used within-subjects design so that each rat served in all 14 conditions. The rats were well trained and apparently adjusted quickly to the prevailing reinforcing stimulus, so that only one session was used in each condition. This is important for it indicates that a much larger design would be feasible, varying other factors such as VI value, amount of solution, and drive level.

The most pertinent aspect of Guttman's report is that his design was a 2 by 7 factorial, and his data exhibited parallelism. Plotted as a function of concentration, the sucrose and glucose curves were approximately parallel. This result seems to support the

joint conclusion that barpress rate was a linear scale, under the given conditions, and that an adding integration rule was at work (Anderson, 1974c, p. 294).

There is, however, a specific objection to the adding model. Incentive value might be expected to be proportional to total sweetness intake, that is, to the product of sugar type and concentration. If response rate in turn was proportional to total sweetness, then the data would obey a multiplying rule and so exhibit a linear fan shape. Under this argument, the observed parallelism would be an artifact of a nonlinear, log response scale.

Some support for the adding interpretation of Guttman's data comes from Clark (1958), who studied effects of food deprivation on press rate of rats under three VI schedules. In terms of the MEV model noted in the Introduction, the deprivation factor should affect the motivation term of the model. The schedule factor should affect the expectancy term, and these two factors should multiply. If press rate is a linear measure, then the data should exhibit the linear fan shape. Indeed, Clark used a linear fan type analysis (his Figure 2) and found surprisingly good support for a multiplying model. His report thus argues against the log response interpretation and for the adding interpretation of Guttman's (1954) data.

Closely related is Dinsmoor's (1952) finding that press rates to interspersed reinforced and nonreinforced discriminative stimuli were mainly proportional across deprivation level. In model terms, the discriminative stimulus would affect the expectancy term and that would multiply with the motivation term.

These reports thus reinforce the hope that VI press rate will be a linear scale. However, the use of other integration tasks seems desirable. In particular, compound adding-multiplying models could provide important information, as noted above.

Guttman's (1954) use of a sensory-type reinforcer deserves notice. The effective incentive was presumably the sweetness taste and, in fact, Guttman related his rat results to human data on taste. This approach has two notable advantages. First, sensory-perceptual integration appears to be a fruitful area for the discovery of algebraic models. Second, cross-species comparisons become feasible because the models provide a validation base for stimulus and response scaling. In particular, direct comparisons between human judgment and animal behavior could become helpful in the problem of the linear response scale.

Blough. A second prototype investigation is Blough's (1972, Experiment 4) study of discrete responses to compound stimuli. Each trial was signaled by illuminating the single response key with a line that varied jointly in wavelength and tilt in a 7 by 7 factorial design. Only one of these 49 stimulus combinations was reinforced, but this reinforcement generalized to other wavelengths and tilts. To insure statistical stability of the more extreme choice proportions, each bird received 850 discrete trials with each stimulus combination.

The main response measure was the normal deviate of the proportion of trials on which the key was pecked within the allowed time.

The theoretical question was whether the response to the compound stimuli could be represented as some simple function of the two component stimulus cues, wavelength and tilt. Blough considered a variety of integration rules suggested by signal detection theory and other sources. His main conclusion, based on a linear fan type of analysis (his Figure 16), was that his data supported a multiplying rule. The factorial plot was not parallel, but showed strong convergence as the two stimulus cues departed from their reinforced levels.

This multiplying rule may seem less plausible than some adding-type rule. For example, an averaging model in which the weight parameter increased with distance from the reinforced level of each cue dimension would give a fairly good account of the data. There is even a possibility for a simple adding model since a log-type transformation on the given response measure would make the data additive.

Blough (1972) rejected this simple adding model because it would require stimulus value to be a nonlinear function of wavelength over the 7-nm range, contrary to a preliminary generalization scaling study. Further, the normal deviate response measure is consistent with work in signal detection. Blough's interpretation seems reasonable, therefore, although by no means conclusive. Change from the response/no response task to a proper two-choice task might help reduce some of the uncertainties over the response scale.

Blough's (1972) report stands out for its mass of discrete-trials data in a fairly large, within-subjects factorial design. Experimental devotion is manifest throughout the article, and many of the problems considered there would be important in further work. Of special interest are the demonstrations of the potential value for study of nonlinear models of using a collateral task for stimulus scaling, and the concern for purification of the response measure in connection with short-latency responses and inattention. Such unglamorous problems of experimental procedure are an organic part of substantive theory.

Farley and Fantino. The last prototype investigation, by Farley and Fantino (in press), also appears to be the first application of functional measurement methods to animal behavior. These investigators studied behavior in a concurrent-chains schedule in which each response key had two independent reinforcement schedules associated with it, one for food and one for shock. The purpose of the study was to test the psychological form of the matching law (see discussion of Equation 6 below) in which the ratio of response rates to the two keys is equal to the ratio of their reinforcement values.

This attempt to test the psychological matching law faces an apparently intractable problem in measurement. The reinforcement value of each key is some integrated

resultant of the associated food and shock schedules. By virtue of this integration, the bird is in effect evaluating the food and shock schedules in a common value currency. These values, however, are not known to the investigator. Even if the values of the separate food and shock schedules could be determined on ratio scales, it would still be necessary to determine the conversion unit between food value and shock value. As Farley and Fantino point out, both of these conditions present difficult problems.

Fortunately, a simple approach is possible. Since the psychological matching law has a multiplying form, it can be tested by means of the linear fan theorem. Accordingly, Farley and Fantino (in press) manipulated the schedules associated with the two keys in a left key by right key factorial design. The data followed the matching law reasonably well, although not perfectly.

A striking feature of this report was the further scaling analysis, which was used to predict behavior in a second experiment. The functional measurement analysis yields a functional scale for the joint food-shock schedules. Farley and Fantino decomposed the scale of the joint schedule into separate food and shock scales, and obtained the conversion unit between these two scales. This analysis employed a special feature of the design, in which the food schedule was constant while the shock schedule varied for the left key, and vice versa for the right key. Two further assumptions, one of linearity and one of food-shock additivity, were also required. With these assumptions, it was possible to predict behavior in a second experiment using considerably different schedule values. Although the two assumptions might not be valid in general, they seemed justified here by the excellent agreement between the predicted and observed behavior in the second experiment.

It may be suggested, however, that the problem of food-shock integration, although important in itself, has two disadvantages for testing the matching law. The first concerns the possible drift over long experiments, which is always a danger despite the careful precautions taken by Farley and Fantino. The second is that only relatively small designs are feasible under such conditions and that may not give sufficient constraint on the response measure, as noted above. Studies like this, that require extended time, might be deferred until firmer evidence on response linearity is available.

Incentive Integration

Incentives are seldom psychologically pure, with only one aspect of value. Foods vary not only in biological nutritive value, but also in various hedonic qualities such as appearance, temperature, texture, and seasoning. Effective incentive value depends on an integration of all these aspects. These integration

processes are quite interesting, for both practical and theoretical reasons.

A more immediate consideration is that some incentive integrations seem likely to obey simple algebraic models. If that can be established, even under restricted conditions, it will provide a valuable opening wedge to the general study of algebraic models of behavior.

One problem is that the integration operation will presumably depend on the relation between the aspects of the incentive. For example, Logan (1960, p. 58) presented parallelism evidence for the tentative conclusion that amount and delay of reward are integrated by adding. That implies that amount and delay have comparable effects, and hence that delay acts as an aversive stimulus *per se*, independent of amount. In utility theory, however, delay is usually considered to act as a weight factor, with a multiplying effect on incentive value. An analogous argument appeared above in the discussion of Guttman's (1954) report. In both cases, a multiplying model could masquerade as an adding model if the response measure followed a logarithmic law of diminishing returns.

To avoid this problem, it seems desirable to seek tasks in which the several aspects of the incentive are psychologically comparable. For example, Klitzner's (1975) study of hedonic integration supports additivity of bitter and sweet taste in humans, and similar results may be attainable in animal studies.

An alternative line of attack would be to use two distinct incentives on each trial, with the hope of establishing an adding rule. To avoid possible interaction between the incentives, they might be of different quality, and perhaps separated a little in space and time. The values of both incentives would be systematically varied over successive blocks of sessions to form an incentive by incentive factorial design. Temporal effects could be handled theoretically in terms of a weight parameter in the serial incentive integration (Anderson, 1974c, Section 6). Additionally, delay of the second incentive could also be varied to form a three-factor design. If delay does act as a weight parameter, then Model 2 of Table 3 is applicable. As noted above, such compound models can be useful in resolving the monotone indeterminacy between adding and multiplying models.

Integration of Drive and Incentive

A prominent question in Hull-Spence theory concerns the integration of drive and incentive. Hull assumed a multiplying operation: $E = H \times D \times K$, where E , H , D , and K denote reaction potential, habit strength, drive, and incentive, respectively. In contrast, Spence assumed that drive and incentive add: $E = H \times (D + K)$.

One experimental test can be obtained by varying H , D , and K in a three-way factorial design. Suppose first that the observed response is a linear scale. Then

the diagnostic properties summarized in Table 3 can be applied directly: Hull's theory corresponds to Model 4, and Spence's theory corresponds to Model 3.

Both Hull and Spence assume that H and D combine by multiplying. This two-way interaction should therefore exhibit the linear fan form. The same holds for the interaction of H and K. These two interactions thus provide a test that at least one theory is correct.

The two theories disagree about drive-incentive integration. The algebraic structure of the models implies that the two-way, D by K data table should form a linear fan according to Hull, a set of parallel curves according to Spence. This factorial graph thus provides a simple test between the two theories.

These diagnostic properties of functional measurement were in substantial part anticipated by Logan (1960, p. 59), Reynolds and Pavlik (1960), Seward (e.g., Seward & Procter, 1960), and Spence (1956, p. 196). Indeed, Seward and Procter used a 2^3 , H by D by K design, although with mixed results.

The present treatment adds three essential elements to the previous work. One is the linear fan analysis. This provides for a proper test of the multiplying relation in terms of bilinear and residual components of the interaction. No less important, the linear fan theorem also provides a basis for stimulus scaling, as demonstrated above.

More fundamental is the treatment of the response scale. As pointed out most clearly by Logan (1960), Hull-Spence theory did not provide adequate methods for obtaining or validating a linear response scale. Logan's own results appeared to support an adding model for drive and both delay and amount of reward. With appropriate caution, Logan emphasized that this interpretation depended on the problematical assumption that his response measure was a linear scale (see also Logan, 1965).

Functional measurement provides a straightforward approach to the case in which the observed response is only a monotone scale. To illustrate, suppose that both theories failed the above tests. It might then be argued that the fault lay in the response scale.

This case can be handled by application of the monotone rescaling procedures outlined above. A simple approach would be to apply a monotone transformation to make the D by K interaction as additive as possible, and test the theories against the other interactions. Spence's theory requires the H by D and H by K interactions to plot as linear fans. In Hull's theory, however, the requisite transformation is logarithmic, which would make all interactions exhibit parallelism. This application provides another illustration of the two-operation logic that was discussed above.

Of course, neither Hull's nor Spence's model may pass the test of goodness of fit. Perhaps the more likely possibility is that drive and incentive are integrated by a compound adding-multiplying rule, or by no simple

algebraic rule at all (see also Grice, 1971; Prokasy, 1967). These complications cannot be pursued here, but present methods provide some leverage on them as well.

The Matching Law

According to the "matching law" (Herrnstein, 1970), an organism will distribute its responses between two alternatives in proportion to their respective rates of reinforcement. This may be written:

$$R_1/(R_1 + R_2) = r_1/(r_1 + r_2), \quad (5a)$$

where R_1 and R_2 are the two response rates and r_1 and r_2 are the two corresponding obtained rates of reinforcement. The mathematically equivalent form,

$$R_1/R_2 = r_1/r_2, \quad (5b)$$

is more useful for certain purposes.

As originally stated, the matching law has the unique property that all four terms are physical observables. That is appropriate to a behavioristic approach. Moreover, it makes the matching law directly testable in terms of the observable data. However, the physical matching law, as it may be called, has met with limited success.

Psychological matching law. There is a growing acceptance of the idea that objective, physical measures of the stimulus values are not theoretically adequate. An extended discussion of this issue and its implications for the behavioristic view has been given by Ebbesen (Note 1) and various comments have been made by others. Instead, the observable reinforcement rates of Equation 5 need to be replaced by subjective values (e.g., Anderson, 1974a, p. 294; Baum, 1974; de Villiers & Herrnstein, 1976; Farley & Fantino, in press; Killeen, 1972; Myers & Myers, 1977; Rachlin, 1971; Shimp, 1975; Ebbesen, Note 1).

In this view, each response alternative would have a "value," denoted by v . Equation 5b would then be rewritten in the form,

$$R_1/R_2 = v_1/v_2 \quad (6)$$

Equations 5b and 6 look alike, but they are fundamentally different. The r values in Equation 5b are observable; the v values in Equation 6 are not observable (Anderson, 1974c, p. 294). Equation 6 may accordingly be called the subjective or psychological matching law.

Nevertheless, almost all investigators have continued to rely on physical measures, either directly, or with some ad hoc transformation. One reason, no doubt, has been suspicion of "subjective" concepts. A more compelling reason has been the lack of a theoretical rationale for stimulus scaling.

Both of these reasons against using subjective stimulus values can be resolved with functional measurement. It deals explicitly with subjective values and provides a straightforward method for seeking them. At the same time, it places the subjective concepts on an objective validation base.

Testing the matching law. Testing the psychological form of the matching law is straightforward (Anderson, 1974c, p. 294). The response measure, R_1/R_2 , will be denoted by R . Equation 6 can then be tested as a multiplying model:

$$R = v_1(v_2)^{-1}. \quad (7)$$

The above linear fan theorem is thus directly applicable.

All that is needed, therefore, is to vary the reinforcement value of the two response alternatives in a two-way factorial design. The observed response data then allow strong tests of the matching law using the simple linear fan analysis that was illustrated in the discussion of Tables 1 and 2.

It deserves emphasis that this approach makes no assumptions about stimulus scaling. The levels of each stimulus factor may be purely nominal. If the matching law is correct, then it should pass the linear fan test. In doing so, it will also provide the stimulus values.

This example illustrates the essential logic of the functional measurement approach. Similar techniques may be applied to the ratio model of Equation 5a and to more complex forms of the matching law that involve multiple stimulus factors or more than two response alternatives. An interesting and important extension to joint shock-reward schedules has been given by Farley and Fantino (in press), as discussed above.

Generalizations of the matching law. Various writers have discussed generalizations of the matching law to allow for multiple determinants of reinforcement besides rate. Baum and Rachlin (1969) considered a generalization of the physical matching law to include amount and delay ("immediacy") of reinforcement: $R = (r_1/r_2)(a_1/a_2)(d_1/d_2)$, where R is the response measure, and r , a , and d denote rate, amount, and delay of reinforcement, respectively.

Killeen (1972) considered a corresponding generalization of the psychological matching law. With a slight change of notation, Killeen's equation may be written: $R = [g(r_1)/g(r_2)] [h(a_1)/h(a_2)] [k(d_1)/k(d_2)]$, where g , h , and k are arbitrary functions, and the other terms have the same meaning as above. Killeen's equation is in the spirit of functional measurement and it avoids the dependence on arbitrary physical measures.

Present methods may be useful in the analysis of these generalized models. To illustrate the method of attack, suppose that two stimulus variables, X and Y , are manipulated for each response alternative in factorial-type design. These variables may include

schedule, amount, delay, or type of reinforcement; and they may include other determinants of response, such as required force. The strength of any response alternative is an integrated resultant of the two stimulus variables and may be denoted by $f(X,Y)$, where f is an unknown function.

The psychological matching law can thus be written: $R = f_1(X_1,Y_1)/f_2(X_2,Y_2)$. The linear fan theorem provides a direct test of this equation, using only the raw data. No scaling of X or Y is needed, and nothing need be known about the f functions. Indeed, $f_1(X_1,Y_1)$ corresponds directly to v_1 above; it is treated as an unknown parameter whose value is to be derived from the linear fan analysis.

If the model passes this preliminary linear fan test, then further study of the integration functions for the separate response alternatives becomes possible. The above model diagnostics apply directly. Thus, Killeen (1972) assumes a multiplying rule with $f_1(X_1,Y_1) = g(X_1)h(Y_1)$. If this is correct, then the two-way, X_1 by Y_1 factorial plot will also obey the linear fan theorem. It seems more generally reasonable, however, to expect an adding-type rule, $f_1(X_1,Y_1) = g(X_1) + h(Y_1)$, and hence that parallelism should obtain.

The g and h functions may be viewed as "psychophysical laws," for they relate the psychological value of the stimulus variable to its physical value. In the present approach, these psychophysical functions are derivative from the more basic psychological law.

A practical problem in the study of these many-variable situations concerns design size. With just two levels of each stimulus variable in the above example, each response alternative corresponds to a 2 by 2 design. The complete design would then be a 2 by 2 by 2 by 2 with 16 experimental conditions. This provides a strong test of the linear fan theorem for the basic ratio model. It also provides a parallelism test of additivity for each separate response alternative. However, it does not provide a test of the assumption of Baum and Rachlin (1969) and of Killeen (1972) that the reinforcement parameters multiply; a 2 by 3 design is the smallest possible size to test the multiplying assumption.

There are various ways by which design size may be reduced, including fractional replication methods developed in statistical design theory. Asymmetrical designs can also be utilized. In the above example, X_1 and Y_1 could be varied in a 2 by 3 design, X_2 and Y_2 in a 2 by 1 design (i.e., with Y_2 held constant). That would provide tests of the basic ratio rule, and of the adding and multiplying rules for the first response alternative.

Bias and undermatching. The psychological form of the matching law also helps clarify the terms "bias" and "undermatching" (Baum, 1974). Bias means that observed response rates are unequal despite equal reinforcement rates. It seems evident, as Baum states,

that bias must reflect some unknown variable that affects preference. Bias represents a failure of the physical matching law, but is allowed for in the psychological matching law. Indeed, the linear fan test remains valid and directly applicable.

Undermatching refers to observed response rates less extreme than predicted by the physical matching law in the absence of bias. Baum treats this as an "error," the reasons for which are "obscure at present." This interpretation reifies a discrepancy from an invalid hypothesis; undermatching merely means that the physical matching law does not apply.

Undermatching can readily be accommodated within the psychological matching law. If the psychological matching law is correct, then undermatching has a simple interpretation. It implies that value is a convex down (negatively accelerated) function of observed reinforcement rate. That would not be surprising since psychophysical studies show that sensations such as loudness and heaviness are convex down functions of physical energy. However, the fact that heaviness is not proportional to gram weight is not properly an error. No more should undermatching be considered as an "error" that requires explanation.

Matching as averaging. Matching can be viewed as the outcome of an averaging process. Each response alternative is assumed to have a weight and a scale value, denoted by w and s , respectively. The weight represents the strength of the tendency, and the scale value denotes its position on the response axis. For the case of two response alternatives with a linear response measure, the expression for the overt response is the weighted average, $R = (w_1 s_1 + w_2 s_2) / (w_1 + w_2)$. Since the response alternatives are discrete, one scale value may be set equal to one, the other equal to zero. The above equation then reduces to the averaging ratio rule, $R = w_1 / (w_1 + w_2)$. More generally, the overt measure of one of a number of discrete response alternatives is equal to its strength divided by the sum of the strengths of all the alternatives.

This averaging interpretation suggests a possible correspondence between behavior and judgment. Averaging processes are pervasive in human judgment, and the averaging model has accounted for a complex pattern of results in some demanding tests. Psychological correspondence between judgmental and behavioral studies is clearly speculative, but may deserve further consideration.

Testability of the matching law. Rachlin (1971) has argued that the matching law in the psychological form of Equation 6 is tautological. If the stimulus values are unobservable, then the investigator has the liberty, if not the obligation, of choosing values that make the model fit the data. Rachlin claimed that this is always possible. In his view, the matching law is true by postulate and "is not an empirical law, but a statement of how reinforcement value is measured" (Rachlin,

1971, p. 250). Similarly, although Killeen (1972) took issue with Rachlin in certain respects, he agreed that the "matching law is indeed a tautology" (p. 490).

Functional measurement makes clear that the matching law is not a tautology. When factorial design is used, the data can fail to obey the matching law. The linear fan analysis provides a strong test.

This capability of functional measurement results from its concern with stimulus integration. When only one stimulus is varied, there are no constraints on the model. However, when two or more variables are varied jointly in factorial design, then adequate constraints are available to provide a rigorous test of the model.

Much the same issue arose in psychophysics. The traditional one-variable approach was unable to provide a validation base for the psychophysical law. Only with a shift to many-variable designs for stimulus integration tasks did a validation base become available (Anderson, 1970, p. 167, 1977b, 202).

This functional measurement approach may be pushed one final step. It is possible to treat the response measure itself as a monotone scale. The model is still testable, as noted above. Ultimately, therefore, only the algebraic structure of the model is at issue.

Nevertheless, Rachlin's (1971) discussion is important, although in a different and more disturbing way than originally intended. Few experiments have used experimental designs or methods of data analysis that are capable of providing a valid test of the psychological form of the matching law. When only a one-way design is used, it is indeed always possible to find a set of stimulus values that will fit the data; that is equivalent to using only one row in the designs of Tables 1 and 2 above. These stimulus values are not tautological, however, but right or wrong.¹

Any behavior law will have limits on its applicability. It is important to develop methods that can assess these limits. The methods currently employed in studies of the matching law are often inadequate and at best inefficient. Functional measurement does not provide an easy cure-all, but it does promise useful assistance.

COMMENT

To balance an undue prominence of scaling considerations, it should be reemphasized that functional measurement is not primarily concerned with measurement but with integration functions. These functions give rise to the need for measurement, but they also provide a base and frame for meeting this need. In this way, measurement is made an organic part of substantive inquiry.

This approach is not novel. Many previous workers in psychology have attempted to base measurement on simple algebraic models (see Anderson, 1970, p. 168, 1974a, p. 291, 1975, p. 480) and the present approach is in the same spirit. The guiding theme has been stated

by Nagle (as cited by Torgerson, 1958, p. 12) that the purpose and justification of measurement is "to establish the equations and theories which are the goal of inquiry."

The present approach does provide a unification of much previous work, both in psychology and in statistics. Also, it has avoided the cul that Torgerson warned against, of preoccupation with mathematical and statistical techniques and neglect of empirical foundations. Useful contributions have been made to a broad range of substantive problems in judgment theory. It is hoped that this approach will also contribute to general behavior theory.

REFERENCE NOTE

1. Ebbesen, E. B. *Operants, scaling, and testing the law of effect*. Unpublished manuscript, 1974. (Available from Department of Psychology C-009; University of California, San Diego, La Jolla, California 92093.)

REFERENCES

- ANDERSON, N. H. An analysis of sequential dependencies. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford University Press, 1959.
- ANDERSON, N. H. On the quantification of Miller's conflict theory. *Psychological Review*, 1962, **69**, 400-414.
- ANDERSON, N. H. An evaluation of stimulus sampling theory: Comments on Prof. Estes' paper. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.
- ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153-170.
- ANDERSON, N. H. Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 2). New York: Academic Press, 1974. (a)
- ANDERSON, N. H. Cognitive algebra. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974. (b)
- ANDERSON, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: W. H. Freeman, 1974. (c)
- ANDERSON, N. H. On the role of context effects in psychophysical judgment. *Psychological Review*, 1975, **82**, 462-482.
- ANDERSON, N. H. How functional measurement can yield validated interval scales of mental quantities. *Journal of Applied Psychology*, 1976, **61**, 677-692.
- ANDERSON, N. H. Failure of additivity in bisection of length. *Perception & Psychophysics*, 1977, **22**, 213-222. (a)
- ANDERSON, N. H. Note on functional measurement and data analysis. *Perception & Psychophysics*, 1977, **21**, 201-215. (b)
- ANDERSON, N. H., & CUNEO, D. O. The height + width rule in young children's judgments of quantity. *Journal of Experimental Psychology: General*, in press.
- ANDERSON, N. H., & FARKAS, A. J. Integration theory applied to models of inequity. *Personality and Social Psychology Bulletin*, 1975, **1**, 588-591.
- ANDERSON, N. H., & SHANTEAU, J. Weak inference with linear models. *Psychological Bulletin*, 1977, **84**, 1155-1170.
- BAUM, W. M. On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, 1974, **22**, 231-242.
- BAUM, W. M., & RACHLIN, H. C. Choice as time allocation. *Journal of the Experimental Analysis of Behavior*, 1969, **12**, 861-874.
- BIRNBAUM, M. H., & VEIT, C. T. Psychophysical measurement: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, 1974, **15**, 7-15.
- BLOUGH, D. S. Recognition by the pigeon of stimuli varying in two dimensions. *Journal of the Experimental Analysis of Behavior*, 1972, **18**, 345-367.
- BOGARTZ, R. S., & WACKWITZ, J. H. Polynomial response scaling and functional measurement. *Journal of Mathematical Psychology*, 1971, **8**, 418-443.
- CLARK, F. C. The effect of deprivation and frequency of reinforcement on variable-interval responding. *Journal of the Experimental Analysis of Behavior*, 1958, **1**, 221-227.
- CLIFF, N. Scaling. *Annual Review of Psychology*, 1973, **24**, 473-506.
- DE LEEUW, J., YOUNG, F. W., & TAKANE, Y. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 1976, **41**, 471-503.
- DE VILLIERS, P. A., & HERRNSTEIN, R. J. Toward a law of response strength. *Psychological Bulletin*, 1976, **83**, 1131-1153.
- DINSMOOR, J. A. The effect of hunger on discriminated responding. *Journal of Abnormal and Social Psychology*, 1952, **47**, 67-72.
- FALMAGNE, J.-C. Random conjoint measurement and loudness summation. *Psychological Review*, 1976, **83**, 65-79.
- FARLEY, J., & FANTINO, E. The symmetrical law of effect and the matching relation in choice behavior. *Journal of the Experimental Analysis of Behavior*, in press.
- GRICE, G. R. A threshold model for drive. In H. H. Kendler and J. T. Spence (Eds.), *Essays in neobehaviorism*. New York: Appleton-Century-Crofts, 1971.
- GUTTMAN, N. Equal-reinforcement values for sucrose and glucose solutions compared with equal-sweetness values. *Journal of Comparative and Physiological Psychology*, 1954, **47**, 358-361.
- HERRNSTEIN, R. J. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 1970, **13**, 243-266.
- HOVLAND, C. I. Clark Leonard Hull. *Psychological Review*, 1952, **59**, 347-350.
- HULL, C. L. Psychology of the scientist: IV. Passages from the "idea books" of Clark L. Hull. *Perceptual and Motor Skills Monograph Supplement* (9-V15), 1962, **15**, 807-882.
- JONES, M. R. From probability learning to sequential processing: A critical review. *Psychological Bulletin*, 1971, **76**, 153-185.
- KILLEEN, P. The matching law. *Journal of the Experimental Analysis of Behavior*, 1972, **17**, 489-495.
- KLITZNER, M. D. Hedonic integration: Test of a linear model. *Perception & Psychophysics*, 1975, **18**, 49-54.
- KLITZNER, M. D. *Small animal fear: An integration-theoretical analysis*. Unpublished doctoral dissertation, University of California, San Diego, 1977.
- KLITZNER, M. D., & ANDERSON, N. H. Motivation $\times$ Expectancy $\times$ Value: A functional measurement approach. *Motivation and Emotion*, 1977, **1**, 347-365.
- KRANTZ, D. H., & TVERSKY, A. A conjoint-measurement analysis of composition rules in psychology. *Psychological Review*, 1971, **78**, 151-169.
- KRUSKAL, J. B. Analysis of factorial experiments by estimating monotone transformation of the data. *Journal of the Royal Statistical Society (B)*, 1965, **27**, 251-263.
- LEON, M., & ANDERSON, N. H. A ratio rule from integration theory applied to inference judgments. *Journal of Experimental Psychology*, 1974, **102**, 27-36.
- LOGAN, F. A. *Incentive*. New Haven: Yale University Press, 1960.
- LOGAN, F. A. Decision making by rats: Delay versus amount of reward. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 1-12.

- LOPES, L. L. Model-based decision and inference in stud poker. *Journal of Experimental Psychology: General*, 1976, **105**, 217-239.
- LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.
- MYERS, J. L. Probability learning and sequence learning. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 3). Hillsdale, N.J.: Lawrence Erlbaum, 1976.
- MYERS, D. L., & MYERS, L. E. Undermatching: A reappraisal of performance on concurrent variable-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, 1977, **25**, 203-214.
- NORMAN, K. L. A solution for weights and scale values in functional measurement. *Psychological Review*, 1976, **83**, 80-84.
- ODEN, G. C. *Semantic constraints and ambiguity resolution*. Unpublished doctoral dissertation, University of California, San Diego, 1974.
- ODEN, G. C. Fuzziness in semantic memory: Choosing exemplars of semantic categories. *Memory & Cognition*, 1977, **5**, 198-204.
- PROKASY, W. F. Do D and H multiply to determine performance in human conditioning? *Psychological Bulletin*, 1967, **67**, 368-377.
- RACHLIN, H. On the tautology of the matching law. *Journal of the Experimental Analysis of Behavior*, 1971, **15**, 249-251.
- REYNOLDS, W. F., & PAVLIK, W. B. Running speed as a function of deprivation period and reward magnitude. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 615-618.
- SEWARD, J. P., & PROCTER, D. M. Performance as a function of drive, reward, and habit-strength. *American Journal of Psychology*, 1960, **73**, 448-453.
- SHANTEAU, J. Component processes in risky decision making. *Journal of Experimental Psychology*, 1974, **103**, 680-691.
- SHANTEAU, J. C., & ANDERSON, N. H. Test of a conflict model for preference judgment. *Journal of Mathematical Psychology*, 1969, **6**, 312-325.
- SHANTEAU, J. C., & ANDERSON, N. H. Integration theory applied to judgments of the value of information. *Journal of Experimental Psychology*, 1972, **92**, 266-275.
- SHIMP, C. P. Perspectives on the behavioral unit: Choice behavior in animals. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 2). Hillsdale, N.J.: Lawrence Erlbaum, 1975.
- SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.
- TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.
- TUKEY, J. W. Analyzing data: Sanctification or detective work? *American Psychologist*, 1969, **24**, 83-91.
- WEISS, D. J. FUNPOT: A FORTRAN program for finding a polynomial transformation to reduce any sources of variance in a factorial design. *Behavioral Science*, 1973, **18**, 150.
- WEISS, D. J., & ANDERSON, N. H. Use of rank order data in functional measurement. *Psychological Bulletin*, 1972, **78**, 64-69.

NOTE

1. This criticism of work on the matching law holds also for the attempts by de Villiers and Herrnstein (1976) to develop

a law of response strength for the standard operant procedure with only one response alternative. This model reads:

$$R = kr/(r + r_e), \quad (8)$$

where R is the observed response measure, r is a measure of reinforcement (e.g., frequency or quantity), r_e is a hypothetical "background level of reinforcement" from extraneous sources, and k is an asymptotic constant. de Villiers and Herrnstein claim that this model does a good job. However, their analysis and conclusions are both open to serious objections.

A primary objection is that de Villiers and Herrnstein do not provide adequate tests of fit. Their main tool is percentage of variance accounted for by the model. These values are of the order of 95% across about 40 different experiments. Although that may seem impressive, higher values can readily be obtained even from models that are badly incorrect (Anderson, 1977b; Anderson & Shanteau, 1977). More seriously, these variance percentages have no relevance or bearing on the model. When R is a monotone function of r, as seemed generally true in their analyses, then some monotone function of r will make the model fit perfectly. These transformed values are the appropriate ones—if the model is correct. But the transformation will also yield a perfect fit even if the model is not correct. By the standard logic of scientific inference, therefore, these variance percentages have zero value as evidence.

de Villiers and Herrnstein (1976) recognized this problem but drew a different conclusion. However, they did present an alternative analysis in terms of testing constancy of k across various conditions. This approach is valid, but the test is weak.

Strong tests of the model become possible when factorial design is used. For example, Guttman (1954) studied response rate in a VI schedule with sucrose and glucose as reinforcers. According to de Villiers and Herrnstein, the estimate of k should be the same for sucrose and glucose, and they found that to be very nearly true. The value of r_e should differ, however, so Equation 8 implies that R for the two sugars should be nearly equal when r is very small, and again when r is very large. For intermediate values of r, however, R should be different for the two sugars. In graphical terms, the factorial plot should have the shape of a slanted barrel, wider in the middle, pinched together at the ends. Numerical calculations with the values of k, r, and r_e used by de Villiers and Herrnstein (1976) show that their model predicts a pronounced barrel shape. The actual factorial plot (Guttman, 1954, Figure 1) is very nearly parallel. What deviations from parallelism are visible are opposite to prediction. As they stand, therefore, the analyses presented by de Villiers and Herrnstein are unsatisfactory and seem to have little bearing on their hypothesis. Similar reanalyses of the other factorial experiments that they consider would be desirable.

(Received for publication May 10, 1977;
revision accepted September 27, 1977.)