Psychological Bulletin
1984, Vol. 95, No. 2, 226-246

Issues in Using Quantitative Rating Scales in Developmental Research

Colleen F. Surber
University of Wisconsin—Madison

Theoretical and practical issues in the use of quantitative judgments in developmental research are reviewed. Judgment is conceptualized as consisting of three components: information valuation, information integration, and translation of impressions into overt responses. It is important to distinguish between developmental changes that occur in different components of the judgment process. Most developmental studies have assumed that rating scales are used in an equal-interval manner, with few attempts to test that assumption. The conditions allowing a conclusion that individuals or groups differ in information integration are outlined, and methods of testing models of the information integration process that do not assume an equal-interval response scale are reviewed. Although these methods require more complicated experimental designs, their benefits in identifying the possible loci of developmental changes in judgments outweigh their inconvenience.

Development can be characterized as involving change in the organization and modes of using information. Modern developmental psychologists study many tasks that require the individual to use multiple sources of information to make various types of judgments. Interest in the way information from multiple sources is used appears to be based on the belief that these processes can reflect fundamental aspects of cognition. For example, Piaget (1965; Inhelder & Piaget, 1958) proposed that there are developmental shifts both in the amount of information children use and in the way the information is used. Piaget's theoretical proposals about developmental change were based largely on patterns of behavior observed in multiattribute tasks (e.g., conservation, perspective taking, moral judgment, and problem solving).

An important current approach to development in such tasks is to collect quantitative judgments and use them to test mathematical representations of the cognitive processes of combining information. The growth in popularity of this approach is consistent with Siegler's (1978) comment that the field of cognitive development "seems to be going toward the increasing use of formalisms to represent what children know" (p. xi).

Applications of mathematical models of judgment in the study of development have involved topics such as moral judgment (Butzin, 1979; Grueneich, 1982; Leon, 1980, 1982a, in press; Surber, 1977, 1982), equity judgments (Anderson & Butzin, 1978), achievement judgments and attributions (Gupta & Singh, 1981; Kun, Parsons, & Ruble, 1974; Surber, 1980), quantity judgments (Anderson & Cuneo, 1978a; Cuneo, 1980, 1982; Leon, 1982b; Verge & Bogartz, 1978), judgments of time, distance, and speed (Wilkening, 1981), and evaluation of likeableness of hypothetical playmates or toys (Butzin & Anderson, 1973; Hendrick, Franz, & Hoving, 1975; Singh, Sidana, & Saluja, 1978). The major aim in such studies is to test competitively models of the way information is combined in making judgments and to describe developmental changes in terms of changes in the parameter values or form of the model. In addition to studies that explicitly test models of information integration, there is the widespread use of quantitative rating scales in developmental studies without critical consideration of implicit assumptions or an explicit attempt to test models of information integration (DePaulo, Jordan, Irvine, & Laser,

Preparation of this article was supported in part by a grant from University of Wisconsin—Madison, the Graduate School.

I am grateful to John R. Surber, Jerome Busemeyer, and an anonymous reviewer for helpful comments.

Requests for reprints should be sent to Colleen F. Surber, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

1982; Heller & Berndt, 1981; Kun, 1977; Shaklee, 1976; Singer & Flavell, 1981; Wellman, Collins, & Glieberman, 1981; Wimmer, Wachter, & Perner, 1982).

This article reviews both theoretical and practical issues in the use of quantitative judgments in developmental research. The first section of the article describes the conditions that will allow an inference of group differences in the process of combining information as opposed to group differences in use of the response scale. The first section outlines the experimental designs and patterns of results that will allow an unambiguous conclusion that there are developmental differences in either the form of the equation describing the way information is combined or the values of the parameters. The issues are illustrated initially with hypothetical examples, followed by an examination of some published research. The concepts presented can also be applied to the problem of inferring differences between individuals or groups that differ in culture, gender, or other ways.

The second part of the article addresses practical problems in analyzing individual patterns of judgment data. This topic is important because group averages may not accurately reflect individual patterns. A particular advantage of algebraic models of judgment is that they can often be applied to the description of individual differences. In order to realize this advantage, however, it is necessary to consider various ways of analyzing the data of individuals. In developmental applications individual data analyses pose special problems because it is difficult to obtain many observations on an individual child.

The central issue of the first part of the article—deciding when observed developmental differences in responses are due only to differences in the use of rating scales or also to differences in the processes of evaluating and combining information—can be regarded as a special case of a more general problem, specifically that of inferring differences in psychological processes or states from between-group (or longitudinally between-times-of-measurement) differences in observed responses. This is a fundamental problem in the study of development that is more often avoided than confronted. The problem in making between-group comparisons of responses can be illustrated by an example. Suppose, on a particular day a researcher in Paris asks a passerby, "What is the temperature today?" The passerby answers, "about 20 degrees." On the same day, a colleague of the researcher asks a passerby in New York the same question and receives the answer, "about 68 degrees." Suppose the researcher then concludes that it was warmer in New York that day than in Paris (or perhaps that New Yorkers feel warmer than Parisians). Of course, degrees celcius is a simple linear function of degrees Fahrenheit, and 20 °C represents the same temperature as 68 °F. Knowing this, it seems likely that the two passersby transformed the subjective temperature into a response using different functions. However, in research there is no a priori way of knowing the function transforming subjective state into observable response and no a priori guarantee that the same response by two groups or individuals represents the same psychological state.

The ultimate solution to this interpretive dilemma, in the author's opinion, is to neither avoid theorizing about psychological processes and states nor "operationally define" responses to be equivalent with psychological states, as is often implicitly done in developmental research. Instead, a hypothesis about the relation between observed responses and psychological processes or states should be embedded in a theory that is richly articulated enough to be capable of predicting a complex network of empirical relations. In the context of such a theory, the hypothesis about the relation between psychological process and response becomes testable rather than a topic for empty argument. The basic idea is that the richer the network of theory and supporting empirical relations, the fewer the alternative explanations, and the greater the inductive support for the theory (Dulany, 1968). The mathematical models of judgment described later constitute such a richly articulated theory. However, the absence of a formal theory does not make a researcher's findings immune to the fundamental issue of how to interpret age-group differences in responses. On the contrary, many findings in the developmental literature can be seen as devoid of scientific value when the between-group equivalence of the psychological processes underlying the observed responses is questioned.

One may reasonably ask, however, what mathematical models of judgment processes can contribute to our understanding of development. First, they can contribute to the precision of our descriptions of developmental changes, what Wohlwill (1973) and others termed the "developmental function." Algebraic models have great descriptive power because they can describe either continuous quantitative developmental changes (as changes in the parameter values) or qualitative developmental changes (as changes in the form of the equation), a longstanding issue in the study of development. With their precise descriptive potential, algebraic models can fulfill one of the important functions of theory in science—specifically, to unify and render understandable a set of phenomena that would otherwise be a miscellaneous collection of empirical relations.

Once a precise, theoretical description of developmental changes in judgment has been established for a given domain, it will be possible for investigators to manipulate variables that may cause such changes and to measure their effects in terms of the model. An example is provided in a study by Leon (in press) in which it was shown that although the form of the equation for combining intent and consequence in judging moral goodness varies considerably among individuals, the combination process used by mothers and their 7-year-old sons tends to match. It is unlikely that this correspondence would have been discovered without a formal model of the information integration process. Leon's findings open the door for developmental researchers to explore the specific aspects of social experience (such as parent–child interactions) that contribute to the development of particular values, a topic that has long been of interest. Leon's study also illustrates the potential of mathematical models of judgment in studying the causes of change in the organization and use of knowledge.

Outline of Judgment

To fully consider the issues involved in describing developmental changes in judgment, an outline of the basic assumptions underlying mathematical models of judgment is needed. A standard view of the judgment process is presented in Figure 1 (Anderson, 1970, 1979; Birnbaum, Parducci, & Gifford, 1971). The first component process is assigning each physical stimulus (ϕ in Figure 1) a psychological value or scale value on the dimension of judgment (s_i in Figure 1). The process of transforming physical stimuli into subjective values is represented as a function labeled H. Anderson referred to this process as "stimulus valuation." The actual stimuli need not have objectively measured physical values, but if they do (as in many studies of perception) the function relating physical value to psychological value is termed the *psychophysical function*. It is important to note that a priori knowledge of the psychological values is unavailable. Instead, psychological values of the stimuli are derived from the equation representing the process of combining information.

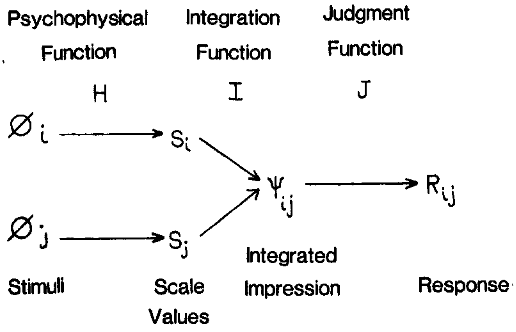

Figure 1. Outline of judgment (from Birnbaum et al., 1971). ϕ_i and ϕ_j at left represent the stimuli that are presented, s_i and s_j represent the subjective values of the stimuli, ψ_{ij} represents the integrated impression of the stimuli, and R_{ij} represents the subject's response on the rating scale.

The second step in judgment is to combine the scale values in some manner to produce an integrated impression of the stimulus combination (ψ_{ij} in Figure 1). Anderson sometimes referred to this as the "psychological law." This process is represented by the information integration function I in Figure 1, which can be an algebraic equation in terms of the scale values of the stimuli and other parameters such as weights. The equation representing the information integration process constitutes a testable theory of the cognitive processes of combining information. Describing developmental changes in information integration is the central concern of many of the studies previously cited.

The integrated impression (ψ_{ij}) of the information cannot be directly observed, however. The observed data consist of responses (R_{ij} in Figure 1) on a rating scale. The process of converting the integrated impressions into responses is represented by the judgment function, labeled J in Figure 1.[1] Anderson (1979) sometimes referred to this process as the "psychomotor law." The judgment function is assumed to be at least a monotonic or rank-order preserving function, although in many studies it is implicitly or explicitly assumed to be linear. Hence, the responses observed in an experiment are theoretically represented as the composition of the three functions: $R = J\{I[H(\phi_i, \phi_j)]\}$. Although it might be possible to express formally the scheme in Figure 1 other than as a mathematical composition of functions (e.g., as a computer program), the mathematical representation is preferred for the present purposes. It should be remembered that the mathematical approach to the cognitive processes of judgment is not necessarily incompatible with other information-processing models.

The outline of the judgment process in Figure 1 is useful in analyzing the possible loci of developmental changes in observed judgment responses. It is apparent that between-group differences in the observed responses could be due to differences in any combinations of the three functions: H, I, or J. The challenge is in separating these three functions to discover the loci of developmental changes. The distinction between I and J in Figure 1 raises the question of how any conclusions at all can be drawn about the form of the integration rule I, if the observed response also reflects the effects of J, the judgment function. In fact, there is a long history of controversy in the field of psychophysics over whether stimulus comparisons involve subjective ratios or subjective differences and over whether subjective values are better measured by magnitude estimation or category-rating procedures (Attneave, 1962; Birnbaum, 1979, 1980; Krantz, 1974; Marks, 1974; Rule & Curtis, 1980; Stevens & Galanter, 1957; Torgerson, 1961). Part of this controversy centers on whether the judgment function is invariant and whether either procedure can justifiably be assumed to have a linear judgment function J (see Birnbaum, 1982a, for an overview). One aim of this article is to describe ways of generating a network of data that is rich enough to permit conclusions about the integration function to be drawn while minimizing the assumptions that must be made about the judgment function.

Implications of Different Assumptions About the Judgment Function

The judgment function, which translates impressions into responses, may change developmentally. A typical strategy in developmental research has been to make assumptions about the judgment function and to use those assumptions to make inferences about the information integration function I. This section examines the implications of different assumptions about the judgment function for inferring developmental changes in either the integration function or psychophysical function for data in which the rank orders are identical across groups.

J Assumed Developmentally Invariant and Linear

The simplest case assumes the judgment function is the same linear function for all ages or subjects in an experiment. Developmental change in the integration function or psychophysical function can then be inferred directly from the observed data. This case is illustrated in Figure 2 for a hypothetical set of data for three age groups. Figure 2 represents hypothetical data from an experiment in which subjects from three age groups judged factorial combinations of Variables A and B. Notice that in Figure 2, Age Group 3 shows a larger effect of both variables than do Age Groups 1 and 2. The larger effect of Variable B can be seen in that the vertical separation of the curves for Age Group 3 is greater than for the

[1] An alternative to the approach of Figure 1 would be to operationally define the integrated psychological impression as equivalent to the observed responses and exclude the judgment function J. To use such an approach would be analogous to using the pre-signal-detection-theory definition of a sensory threshold in choice experiments. That is, the judgment function in Figure 1 can be regarded as analogous to the theoretical construct of the response criterion in signal detection theory. To ignore the judgment function in a study using quantitative ratings would be like ignoring the response criterion in a choice experiment.

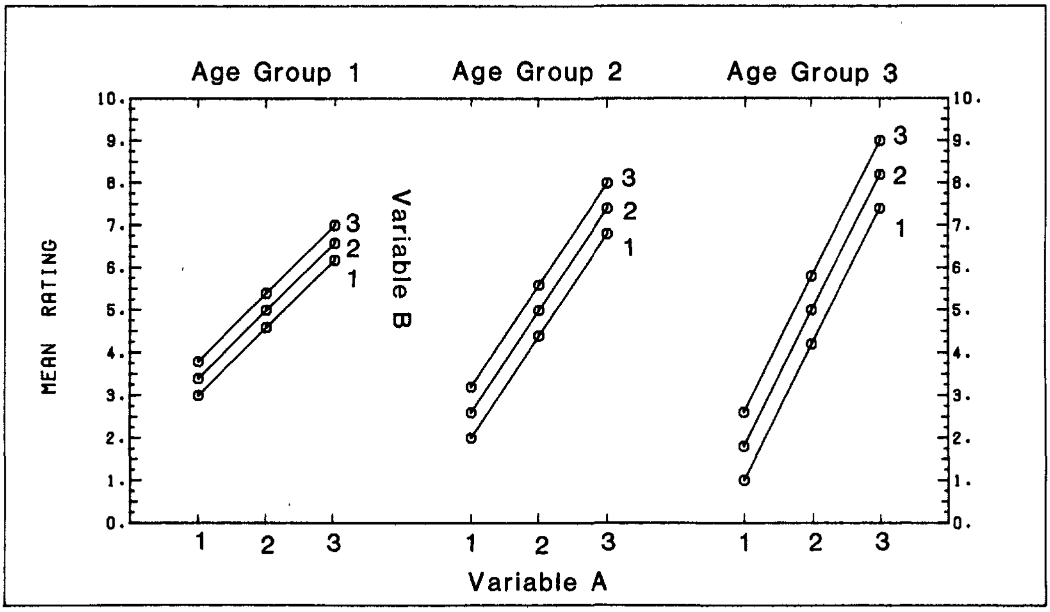

Figure 2. Hypothetical data for judgments of factorial combinations of three levels of Variable A (abscissa) and Variable B (separate curves).

other age groups, and the larger effect of Variable A can be seen in the differences in slope. Under the assumption that the judgment function is developmentally stable, the developmental differences in the size of the main effects can be inferred to be due to changes either in some aspect of the integration function (e.g., weights of variables) or in the scale values of the stimuli (i.e., for Age Group 3 the manipulation of the variables may have seemed more extreme; thus, the range of scale values may have been wider). Such conclusions are sometimes drawn (cf. Miller, 1982), but without explicit acknowledgment of the assumption that the judgment function is developmentally invariant.

A different set of hypothetical data is given in Figure 3. Under the assumption that the judgment function is linear and developmentally invariant, the age differences in Figure 3 would be attributed to the information integration function *I*. Changes in the scale values

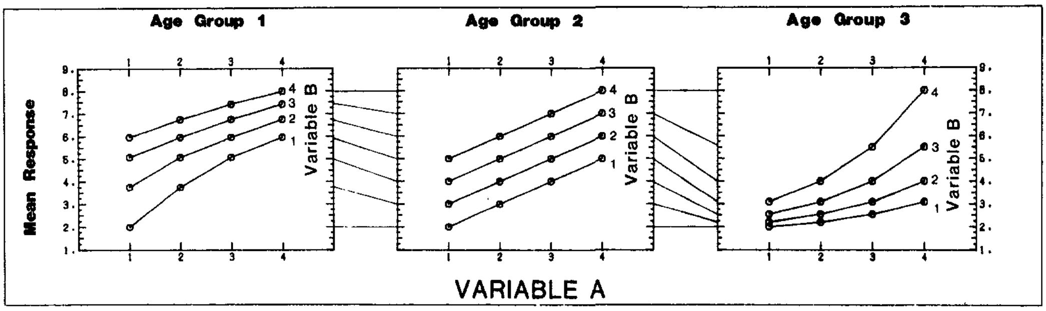

Figure 3. Hypothetical results illustrating that differences in response patterns do not necessarily imply differences in information integration. (The three panels represent responses based on an additive integration function, subjected to a log transformation [left-hand panel], no transformation [middle panel], and an exponential transformation [right-hand panel]. The rank orders of the data are identical in all three panels. Lines connecting the panels show the transformations.)

are also possible but could not totally account for the observed differences in the response patterns. The results for Age Group 2 can be represented by an additive integration process because the data form a parallel set of curves. The results of Age Groups 1 and 3 can be represented by any of several nonadditive integration processes (e.g., differentially weighted averaging or multiplying).

J Assumed Linear but Developmentally Variable

Relaxing the assumption of developmental invariance in the judgment function, even while retaining the assumption of linearity, produces different conclusions for the hypothetical data of Figure 2. Assuming that the judgment function is a different linear function for each age group allows the data of Figure 2 to be described without any developmental changes in either the integration function or the psychophysical function. It can be shown that this is the case simply by linearly transforming the data of each panel to the same range. The data of the three panels will then be identical and can be represented by an additive integration function with developmentally constant parameters. Under these assumptions, a researcher might conclude that there is no change in the way information is evaluated or combined. To predict the developmental change in Figure 2, one need only assume that the age groups differ in their tendency to spread their responses over the full length of the response scale, a phenomenon that could occur for a variety of reasons. One possibility is that there may be more random responding by the younger children than by the older children. Averaging random responses with a mean at the midpoint of the response scale would have the effect of pulling the mean responses toward the midpoint of scale, approximately as shown in Figure 2.

For the hypothetical data of Figure 3, the conclusions above regarding developmental change in the integration function would not be altered by allowing the judgment functions to be linear but developmentally variable. Any linear transformation of the ordinate can be applied to the means of Figure 3 without distorting the patterns of the mean judgments.

J Assumed Developmentally Variable and Not Necessarily Linear

Interpreting the results of any experiment using quantitative ratings becomes considerably more complex if the judgment function J is not assumed to be linear but instead is assumed only to be monotonic. This has been noted previously by many authors (Anderson, 1977; Birnbaum, 1974a; Bogartz, 1976; Bogartz & Wackwitz, 1970), but its consequences for conclusions in developmental research need to be stated explicitly. In Figure 3, the data patterns of the three age groups can all be derived from an additive integration function combined with either a negatively accelerated judgment function (Age Group 1), a linear judgment function (Age Group 2), or a positively accelerated judgment function (Age Group 3). The rank orders of the means are identical in the three panels, and the transformations of the ordinate mapping one data pattern onto another are shown by the lines connecting the panels. (The left and right panels of Figure 3 were generated by applying logarithmic and exponential transformations, respectively, to the additive data pattern.)

The example in Figure 3 illustrates two important problems in interpreting developmental differences when the shape of the judgment function is unknown. First, it is possible for investigators to conclude erroneously that there is a developmental change in information integration when in fact there is none. Second, it is also possible for investigators to conclude erroneously that there is no developmental change in information integration when in fact there is. An investigator who obtained data like those in Figure 3 might conclude that the information integration process can be represented by an additive function for all three ages. Of course, in the absence of knowledge about the judgment function J, it is also possible that the information integration process does change.

How likely is it that drastically different judgment functions would be encountered in research? Research with both social and nonsocial stimuli has shown that manipulation of the stimulus distribution can change the pattern of responses from a converging interaction such as that shown in the left-hand panel of

Figure 3 to a diverging interaction such as is shown in the right-hand panel (Birnbaum et al., 1971; Mellers & Birnbaum, 1982, 1983). Birnbaum and his colleagues concluded that changes in the judgment function could best account for the shifts in the interaction patterns caused by changes in the stimulus distribution. Given the uncertainty regarding the shape of the judgment function, it is obviously desirable for conclusions about the integration function to depend only on the assumption that the judgment function is monotonic.

Table 1 presents a summary of relevant literature on children's information integration including the topics studied, the authors' conclusions regarding the information integration function, the authors' apparent assumptions about the form of the judgment function, and whether there was any experimental attempt to rule out nonlinearity in the judgment function as an explanation of the results. A symbol is placed next to each integration function for which the conclusion depends on assuming a linear judgment function. It is clear from Table 1 that it is routine for investigators to assume that the judgment function is linear, and it is rare for investigators to attempt to test the assumption that the judgment function is linear.

Constraints on Interpreting Judgment Data

The hypothetical examples in Figures 2 and 3 were constructed specifically to illustrate the ambiguities that arise when the rank orders of the data are identical across age groups. When the rank orders of the data are not identical across ages (assuming that the rank order differences are not due to error), age differences in either information valuation H or information integration I can be inferred without assuming that the judgment function is either linear or developmentally stable. This section describes (a) the way in which the rank order characteristics of a set of data can help determine the nature of the integration function I and (b) experimental designs and methods for separating assumptions about the judgment function J from the integration function I. Wherever possible, the principles are illustrated with actual research findings.

Rank Order Constraints

Most developmental research has used simple two-variable factorial designs to study information integration. Because of this, the rank order predictions of several common models of information integration for two-factor designs are described in the following section. Unfortunately, the rank order properties will not always help in determining the form of the information integration function for experiments using only two variables. No attempt is made here to deal with the issue of measurement error in rank order methods. Other discussions of rank order methods are available elsewhere (Anderson, 1977; Birnbaum, 1982b; Busemeyer, 1980; Krantz & Tversky, 1971; Weiss & Anderson, 1972).

Multiplying versus additive models. Some of the most interesting and important research on algebraic models of children's judgments has focused on the issue of when or whether there is a transition from an additive integration process to a multiplicative one (Anderson & Cuneo, 1978a, 1978b; Bogartz, 1978; Cuneo, 1982; Kun et al., 1974; Surber, 1980; Verge & Bogartz, 1978; Wilkening, 1979, 1981). Distinguishing between these models without assuming the judgment function to be linear would be desirable. Unfortunately, any set of data that is ordinally consistent with an additive model is ordinally consistent with a multiplicative model. Also, unless the subjective values of one of the variables include zero or are negative, any set of data that is ordinally consistent with a multiplying model is also ordinally consistent with an additive model. This fact can be easily appreciated by noting that a log transformation converts a multiplying model with positive scale values to an additive model. Thus, the additive and multiplying models cannot usually be distinguished using ordinal information. Because of this, it is usually not possible to infer a developmental shift from adding to multiplying based on ordinal characteristics of the data.

The exchange between Anderson and Cuneo (1978a, 1978b) and Bogartz (1978) illustrates the difficulty in distinguishing additive and multiplying models. Anderson and Cuneo (as well as Wilkening, 1979) concluded that there is a developmental progression in judging area from an additive to a multiplying function. Bogartz (1978) questioned their conclusion, citing the data of Verge and Bogartz (1978) and raising the issue of how the response scale used might influence the conclusions drawn

Table 1
Summary of Information Integration Literature Using Children as Subjects

Study	Topic	Author conclusion for I	Assumed J	Attempt to determine whether J linear
Anderson & Butzin (1978)	Equity	Relative ratio[a]	Linear	None
Anderson & Cuneo (1978a)	Area Liquid quantity	Additive[a] Height only	Linear	Aggregation design (Experiment 6)
Butzin & Anderson (1973)	Liking	Averaging	Linear	None
Cuneo (1980)	Area Liquid quantity	Additive[a] Height only	Linear	None
Cuneo (1982)	Numerosity	Additive[a] Multiplicative[a]	Linear	Change in stimulus values changed response pattern
Grueneich (1982)	Morality	Linear (additive or averaging)[a] Consequence only Intention only	Linear	None
Gupta & Singh (1981)	Achievement	Averaging	Linear	None
Hendrick et al. (1975)	Liking	Averaging	Linear	None
Kun et al. (1974)	Achievement	Additive[a] Multiplicative[a]	Linear	None
Leon (1980)	Morality	Averaging Configural Multiplicative[a] Consequence only Intent only	Linear	None
Leon (1982b)	Area	Multiplicative[a] Visual extent[a] Proportionality	Linear	None
Leon (1982a)	Morality	Additive[a] Configural	Linear	None
Leon (in press)	Morality	Additive[a] Configural Consequence only	Linear	None
Miller (1982)	Achievement	Additive[a] Interactive[a]	Linear, but author urges caution	None
Singh et al. (1978)	Liking	Averaging	Linear	None
Surber (1977)	Morality	Averaging	Linear, but author urges caution	None
Surber (1980)	Achievement	Averaging with configural weights	Monotonic	Scale-free test
Surber (1982)	Morality	Averaging	Monotonic	None
Verge & Bogartz (1978)	Area	Multiplicative[a] Height only Width only	Linear	None
Wilkening (1979)	Area	Additive[a] Multiplicative[a]	Linear	None
Wilkening (1981)	Velocity, time, distance	Additive[a] Multiplicative[a] Distance only	Linear	None

[a] This conclusion is contingent on assuming the judgment function J to be linear.

by an investigator (as well as raising issues of statistical power and experimental design). Anderson and Cuneo (1978b) replied, in part, that Verge and Bogartz's response scale (an area-matching task) may have produced the diverging fan pattern of data (and the conclusion that height and width combine multiplicatively) when the subjects actually combined the information additively. The purpose of the present discussion is not to decide who was correct in this instance but to point out that these investigators were arguing partly over who had procedures that would yield a linear judgment function. The data of both were ordinally consistent with either an additive or multiplicative integration function.

Adding versus averaging models. Adding and averaging models of information integration cannot be distinguished in a two-variable factorial design unless each type of information is also presented separately. An experiment by Butzin and Anderson (1973) illustrates the ordinal test between the adding and averaging models. Butzin and Anderson asked children to judge their liking for playing with pairs of toys that varied in desirability. An averaging model for this task can be written

$$R_{ij} = J[(w_1 s_{1i} + w_2 s_{2j} + w_0 s_0)/(w_1 + w_2 + w_0)],$$

where w_1 and w_2 are the weights given to the toys presented first and second within each pair, s_{1i} and s_{2j} are the psychological values of the ith and jth toy from the first and second set respectively, the term $w_0 s_0$ represents the weight and scale value of the subject's impression in the absence of any information, and J is a monotonic function. The subject was also asked to judge the desirability of a single toy (instead of a pair); the averaging model for this task can be written

$$R_i = J[(w_1 s_{1i} + w_0 s_0)/(w_1 + w_0)],$$

where the terms are interpreted as above, and J is the same monotonic function as when two toys are presented. Because the weight of information not presented, w_2, is omitted from the denominator, the averaging model predicts that the effect of the first set of toys will be greater when they are presented alone than when combined with a second toy. If the results are graphed, the curve for the toys presented alone should be steeper and should cross one or more of the curves for the pairs of toys combined. This is the result obtained by Butzin and Anderson (1973) for two age groups.

In contrast to the averaging model, the adding model predicts that the net effect of information presented alone should be equal to its effect when combined with other information. Since the additive model has no denominator term, the effect of each variable should be independent of the other information presented with it. When graphed, the curve for the toys presented alone should not cross the curves for the pairs of toys. When the crossover predicted by the averaging model is obtained, there is no monotonic transformation of the response scale that will allow the additive model to account for the data. Thus, the additive and averaging models can be distinguished by the rank order characteristics of the data if each type of information is presented separately as well as in factorial combination. This means that a developmental change from adding to averaging can be detected without assuming the judgment function to be linear or developmentally stable. Such a developmental shift has not been obtained to date. Instead, researchers who test the averaging model typically find that it applies to all age groups studied.

Inferring change in weight or scale value. One important goal of many developmental studies using quantitative responses is to discover whether there are changes in the relative effects of different types of information. For the hypothetical example in Figure 2, in which there are no rank order differences across the age groups, any conclusions regarding age change in the impact of the variables are implicitly based on the assumption that the judgment function J is developmentally invariant.

An experiment by Surber (1977), intended to assess developmental change in the relative impact of intent and consequence information on moral judgments, illustrates how rank order changes can allow the investigator to exclude the judgment function as the source of the developmental differences. Table 2 presents the means and rank orders of the means from Surber's Experiment 1. Table 2 shows that there are age differences in the rank orders. If the integration function is assumed to be either an averaging or adding model, these changes in rank order can be interpreted in two ways: (a) the weight of either intent or consequences may change with age or (b) the range of scale values of either intent or consequences may change with age. Because the models predict that the net effect of a variable is a function of its weight and the range of scale values, either or both of these changes can produce the obtained rank order changes. Because of the obtained rank order changes with age, however, the developmental changes cannot be attributed solely to the judgment function *J*.

Separating the scale value and weight interpretations of change in the effect of a variable is important in developmental work. A developmental change in scale value represents a change in the meaning or interpretation of a social stimulus, whereas a developmental change in weight represents a change in the subjective importance of a stimulus. These two interpretations can be thought of as roughly analogous to Flavell, Beach, and Chinsky's (1966) distinction between production and mediation deficiency in learning tasks. Scale value changes in social judgments can be interpreted as changes in understanding, whereas weight changes represent changes in the use of what one understands.

Weight and scale value cannot be separated if information is combined additively (Schonemann, Cafferty, & Rotton, 1973). However, if the information is combined according to an averaging model, it is possible to separate weight from scale value (Anderson, 1973; Norman, 1976, 1981). A design used by Surber (1982) illustrates a method for separating the weight and scale value interpretations of developmental change. In this experiment, the stimuli were motives combined with consequences or with other motives. An averaging model for this task can be written

$$R_{ij} = J[(w_1s_{1i} + w_2s_{2j} + w_0s_0)/(w_1 + w_2 + w_0)],$$

where w_1 and w_2 represent the weights assigned to a given type or source of information, s_{1i} and s_{2j} represent the scale values of the *i*th and *j*th stimuli of the first and second types,

Table 2
Mean Judged Naughtiness and Rank Orders of the Means for Experiment 1 of Surber (1977)

		Kindergarten children		2nd-grade children		5th-grade children		Adults	
Intention	Consequence	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
Low	Low	1.16 (0.125)	1	1.15 (0.069)	1	1.67 (0.260)	1	1.89 (0.351)	1
	Medium	2.16 (0.269)	3	2.51 (0.180)	3	2.30 (0.248)	2	2.33 (0.373)	2
	High	4.44 (0.490)	7	4.31 (0.250)	6	2.79 (0.178)	3.5	2.44 (0.412)	3
Medium	Low	1.44 (0.183)	2	2.05 (0.191)	2	2.79 (0.281)	3.5	4.56 (0.530)	4.5
	Medium	2.64 (0.355)	5	3.33 (0.199)	5	2.85 (0.200)	5	4.56 (0.338)	4.5
	High	4.80 (0.444)	8	4.54 (0.275)	8	3.73 (0.227)	6	4.89 (0.423)	6
High	Low	2.36 (0.439)	4	3.10 (0.324)	4	4.21 (0.356)	7	6.11 (0.309)	7
	Medium	2.84 (0.467)	6	4.46 (0.289)	7	4.67 (0.274)	8	6.22 (0.324)	8
	High	5.68 (0.479)	9	5.49 (0.214)	9	5.21 (0.313)	9	6.33 (0.333)	9

Note. Values in parentheses are standard errors of the means. *N*s are 25, 39, 33, and 9 for kindergarten children, 2nd- and 5th-grade children, and adults, respectively.

respectively, w_0s_0 represents the impression in the absence of any information, and J is a monotonic function. Assume that w_1 is the weight of motive information when it is presented first. The weight given to the second piece of information, w_2, depends on whether it is a consequence or motive.

According to the averaging model, the net effect of motives when presented first should be a linear function of $w_1/(w_1 + w_2 + w_0)$, and so should depend on whether consequences or other motives are the second type of information. This is the case because the net effect of motives when presented first depends in part on the value of w_2 in the denominator. If motives presented first have a larger effect when combined with consequences than when combined with other motives, then the weight of consequences must be less than the weight of motives (i.e., w_2 has a higher value when it represents motives than when it represents consequences). Nothing need be assumed about the scale values except that they are constant across conditions within each age group. Using this logic, Surber (1982) determined the relative ordering of the weights given to motives and consequences. There was a developmental shift from weighting consequences more than motives to weighting motives more than consequences. The type of differences predicted when weights change are ordinal changes and cannot be accounted for by change in either the range of the scale values or the judgment function.

One general strategy for separating weight and scale value is to combine factorially one variable with each of two others (e.g., A × B and A × C factorial designs) under conditions in which it is plausible to assume that the scale values of Variable A and its weight term are independent of whether it is combined with Variable B or Variable C. This design is the minimum that will allow the averaging model to be tested and the ordering of the weights of Variables B and C to be determined (Norman, 1976, 1981). These methods can be used to separate the weight and scale value interpretations of other developmental changes in the relative impact of information. For example, Gupta and Singh (1981) found that 6- and 7-year-olds were more likely to make use of motivation than ability in predicting performance. This result (as well as similar findings by Kun et al., 1974) suggests that the weight of motivation may be higher than the weight of ability, but that this relation may reverse as people grow older. Analogously, based on the effect sizes, Miller (1982) proposed that the weight of noise level compared with interest level in predicting learning increases with age. Both of these hypotheses about developmental change in weight could be tested more definitively by combining the variables factorially with a third variable as outlined previously.

Separation of the influence of weight and scale value on the net effect of a variable is important beyond the domain of developmental psychology. A well-known problem in research attempting to discover the relative importance of variables is that measures such as the size of an effect or the proportion of variance accounted for depend on the particular values with which a variable is instantiated. The separation of weight and scale value provides a method for measuring the importance of variables that does not depend on the particular stimuli chosen by the experimenter.

Two-Operation Experiments

A powerful approach to the problem of determining whether an observed developmental change is due to the judgment function as opposed to the information integration function is to embed the integration process of interest in another judgment task (Anderson, 1971; Birnbaum, 1974a, 1978). The basic rationale of two-operation integration experiments is that by obtaining a richer network of data, the number of interpretations that are compatible with the complete pattern of data will be reduced. There are two general types of two-operation designs used, which will be termed aggregation designs and comparison designs. In aggregation designs, the subject is asked to judge the aggregation (e.g., total intensity) of two stimuli, each composed of components. In comparison designs, the subject is asked to compare two stimuli, each composed of components.

Aggregation designs. The best example of a stimulus aggregation design is found in Anderson and Cuneo's (1978a) Experiment 6. In this experiment, children were asked to judge the combined area of two rectangles at a time (e.g., "How much is there to eat in both rectangles?"). Children were shown combinations of rectangles generated by a 4 (7 × 7, 7 × 11, 11 × 7, and 11 × 11 cm) × 2 (6 × 5 and 10 × 8 cm) factorial design. Anderson and Cuneo hypothesized that combined area should be judged according to an additive combination rule, $R_{12} = J(\psi_1 + \psi_2)$, where R_{12} is the response to the two rectangles, J is any monotonic function, and ψ_1 and ψ_2 represent the impressions of the area of each rectangle. Anderson and Cuneo's goal was to decompose ψ_1 and ψ_2 in order to discover how width and height are combined in forming an impression of the area of a single rectangle.

One approach to decomposing the total area judgments is to test the metric fit of the additive model for judgments of total area using the significance test of the interaction in an analysis of variance. If the interaction is nonsignificant (given reasonable statistical power), then the additive model can be retained, and the judgment function for total area is assumed to be approximately linear. Under these assumptions, the marginal means from the factorial design provide an estimate of the ψ values for the rectangles. Anderson and Cuneo used this method to estimate the values of the 7 × 7, 7 × 11, 11 × 7, and 11 × 11 cm stimuli. Because these stimuli form a height-by-width factorial design, the factorial plot and statistical tests of the values of these four stimuli can be used to test the hypothesis that height and width are combined using a multiplying as opposed to an adding rule. Based on the results of these tests, Anderson and Cuneo concluded that 5-year-olds combine height and width additively in judging area.

The aggregation judgments can also be decomposed by assuming that J is only monotonic, followed by a transformation of the judgments to fit the additive model using a procedure such as MONANOVA (Kruskal & Carmone, 1969). The marginal means of the transformed values can then be used to test between additive and nonadditive integration processes for the components. This approach does not require the assumption that the judgment function is linear, but it assumes the validity of the additive model if there are no ordinal violations of additivity. A discussion of problems in monotonic transformations of response scales is found in Busemeyer (1980).

One drawback of the use of aggregation models is that the rank order characteristics of the aggregation ratings usually cannot discriminate between additive and nonadditive integration of the components. The aggregation approach represents a significant step forward, however, because it can add converging evidence to results obtained from separate judgments of each stimulus.

Comparison designs. Comparison designs have not received much attention in developmental research as a way of separating developmental change in the integration and judgment functions. This is the case in spite of the fact that children are frequently asked to make choices between items, as in much of Piaget's work and other developmental research (Siegler, 1976; Siegler & Richards, in press). Birnbaum (1978, 1982a) has emphasized the importance of comparison judgment tasks for separating the integration and judgment functions in psychophysics. In the comparison method, the subject can be asked to judge the difference between two stimuli, each of which is composed of two attributes. The judged difference is hypothesized to follow the subtractive model: $R_{12} = J(\psi_1 - \psi_2)$, where R_{12} is the judged difference between Stimulus Compounds 1 and 2, J is a monotonic function, and ψ_1 and ψ_2 are the impressions of the two stimulus compounds. For example, in judgments of area, children could be presented with pairs of rectangles and asked, "Who would have more to eat, the person with cookie 1 or the person with cookie 2?" Following an initial decision, the child can be asked to make a quantitative rating of how much more. The difference ratings can then be used to test the integration process for height and width using what Birnbaum (1974a; Birnbaum & Veit, 1974) called the "scale-free" method. Such an experiment could add support to Anderson and Cuneo's (1978a) conclusion that there is developmental change in combination of height and width.

The scale-free method is illustrated in a developmental study by Surber (1980) in which children judged performance, given ability and effort information. Kun et al. (1974) had previously concluded that there is a developmental shift from additive to multiplicative combination of ability and effort. Surber collected judgments of differences in performance in order to use the scale-free method to distinguish between these models of information integration. An additive model for judgment of task performance based on ability and effort can be written $R_{ij} = J(w_A s_{Ai} + w_E s_{Ej})$, where w_A and w_E are weights that are constant over all levels of s_{Ai} and s_{Ej}, the s_{Ai}, and s_{Ej} are the scale values of the ith and jth levels of ability and effort, and J is a monotonic function. When subjects are asked to judge the differences between pairs of stimuli, the additive model predicts that for pairs of stimuli involving equal ability information, the judgment should depend only on the difference in effort:

$$R_d = [(w_A s_{Ai} + w_E s_{Ej}) - (w_A s_{Ai} + w_E s_{Ek})]$$
$$= J[w_E(s_{Ej} - s_{Ek})].$$

An analogous conclusion holds for pairs of stimuli differing only in ability.

In contrast, the multiplicative model predicts that the judged difference will depend on both the difference in effort and the level of ability:

$$R_d = J[(s_{Ai} \times s_{Ej}) - (s_{Ai} \times s_{Ek})]$$
$$= J[s_{Ai}(s_{Ej} - s_{Ek})].$$

Thus, the judged difference should vary directly with the value of ability as well as with the difference in effort. The subtractive model for difference judgments can be tested either ordinally or metrically, as described for the additive model. The values obtained from the difference judgments can be tested for evidence of an additive versus nonadditive combination of ability and effort. Surber (1980) concluded that the difference judgments supported the multiplying combination of ability and effort for 5th- and 6th-grade children and adults. Although the experiment was not designed to provide a full test of the subtractive model, there were no ordinal violations of it in the judgments of differences in performance.

An advantage of the use of difference judgments is that additive versus nonadditive integration of the dimensions of each stimulus makes different ordinal predictions for judgments of differences (Birnbaum, 1974a). Thus, difference judgments can provide an ordinal test of the additive integration rule for the component dimensions (e.g., ability and effort or height and width) without assuming a linear judgment function.

Assumptions in two-operation experiments. One assumption that has already been mentioned is that the aggregation or comparison follows a particular hypothesized model (for the previous examples, either an additive or subtractive model). This assumption can be tested ordinally, however, so that the method does not require assuming a linear judgment function.

Two-operation experiments also assume that the information integration process of primary interest (e.g., the combination of height and width to estimate area or the combination of ability and effort to estimate performance) is not disturbed by embedding it in a task requiring either aggregation or comparison. This assumption is obviously open to question in developmental research. When presented with more complex cognitive tasks, children may ignore some of the information (Anderson & Butzin, 1978) or may change to a strategy that is more completely mastered (Shatz, 1978). Thus, it is possible for the results of a two-operation experiment to contradict those of a one-operation information integration experiment and not provide an answer to the original research question.

When subjects do change strategies in making judgments in a two-operation task, it seems likely that they would either (a) centrate on one dimension of both stimuli across all trials or (b) cancel a stimulus dimension that has equal value across the two stimuli of a trial. For example, subjects might judge the total area of two rectangles by attending only to height (centration). If this were the case, the additive model would fit the total area judgments, but the derived ψ values would show an effect of only height. Thus, centration should be obvious in the data, and an investigator obtaining such a pattern would undoubtedly not attempt to generalize across the one- and two-operation judgments.

The strategy of canceling or ignoring a dimension that has equal value for a stimulus pair seems most likely to occur when comparing stimuli. For example, given a 7 × 11 and a 7 × 7 cm stimulus, the subject may judge the difference in area by canceling the equal-valued dimension and relying only on the dimension for which the values differ. The dimension that differs will vary between stimulus pairs. For the difference judgment task, this strategy will produce data that agree with the subtractive model. The ψ values derived will show an additive pattern, however. Under these conditions, the cancellation strategy could lead an investigator to the erroneous conclusion that an additive integration function holds. If the judgments in a one-operation task show significant interactions (e.g., height by width or ability by effort), but the impression values derived from difference judgments support the additive model, the investigator should be cautious about concluding that the two dimensions are combined additively. Thus, difference judgments provide the most leverage when they support a nonadditive integration process also found in a one-operation experiment, as in the experiment of Surber (1980).

Scale Convergence Criterion

The integration function and judgment function can sometimes be separated by using more than one type of judgment task and assuming that the scale values for the stimuli are independent of the judgment task. The rationale behind the scale convergence criterion is that measured psychological values should have at least some generality across tasks if they are to have any predictive power or be theoretically useful (Birnbaum, 1982c). In contrast, Marks (1982) assumed that psychological values vary with the experimental procedure (e.g., magnitude estimation vs. category rating).

The scale convergence criterion is illustrated in an experiment by Birnbaum and Veit (1974). Adults were asked to judge both the differences and ratios of weights lifted simultaneously in the left and right hands. They initially hypothesized that the two types of judgments would be based on subtractive and ratio integration processes, respectively, with linear judgment functions. The analysis of variance showed the expected results: Instructions to judge differences resulted in an approximately parallel set of curves; instructions to judge ratios resulted in an approximately bilinear fan of curves.

The scale convergence criterion required that the scale values derived from the fit of the integration functions to the two sets of judgments be linearly related. This was not the case in Birnbaum and Veit's (1974) experiment, however. Instead, the scale values of heaviness from the subtractive and ratio models showed a curvilinear relationship. Birnbaum and Veit concluded that rather than using two different integration operations when instructed to judge ratios and differences, subjects used the same integration process and scale values but different judgment functions. This conclusion was based on the finding that the scale values of heaviness were linearly related when the data from the ratio judgment task were monotonically transformed to fit the subtractive model. An analogous conclusion was drawn by Krantz (1974) for judgments of differences and ratios of brightness. Thus, the assumption of scale convergence provides a criterion for distinguishing between the information integration process and the judgment function.

There appears to be no example of developmental research in which the scale convergence criterion has been employed. One problem is that in order to evaluate scale convergence it is desirable to have a relatively large number of stimulus values. For example, Birnbaum and Veit (1974) used seven stimulus values, requiring 49 values in the factorial designs for each type of judgment. In developmental work, either the willingness or ability of the subjects to make a large number of judgments is questionable. However, with a sufficient number of practice trials it may be reasonable to split an experimental design across two or more sessions (see Surber, 1982, for an example). The scale convergence criterion is also easiest to apply in psychophysics, where the subject can be instructed to make two different types of judgments.

How can the scale convergence criterion help in studying developmental change in information integration when the experimenter does not wish to attempt to influence that process with instructions? An example for which scale convergence would have been relevant is Wilkening's (1981) study of the development of velocity, time, and distance concepts. In his experiment subjects judged distance traveled (given time and velocity), velocity (given distance and time), and time (given distance and velocity). Wilkening concluded that judged distance was a multiplicative function of time and velocity, for 5-year-olds to adults. For judgments of time, he concluded that there was a developmental shift from a subtractive integration process for 5-year-olds (time = distance − velocity) to the ratio integration process for 10-year-olds and adults (time = distance/velocity). For judgments of velocity, Wilkening concluded that a subtractive process held for adults and 10-year-olds (velocity = distance − time), whereas 5-year-olds appeared to ignore time information. These conclusions were based on the implicit assumption that all the judgment functions are linear.

If Wilkening (1981) had used more stimulus levels (the design for all three types of judgments was a 3 × 3 factorial), the scale convergence criterion could, in principle, be used to add leverage to the claims of developmental changes in the processes of combining information. Specifically, the scale values of the stimuli could be compared across judged dimensions. If there is really a developmental shift from the subtractive to ratio integration function, then when these models are fit to the data the scale values should correspond across judged dimensions. The scale values for velocity obtained from judgments of distance and time, for example, could be tested for linearity at each age level. If Wilkening's conclusion about the developmental change in the integration function for judging time is correct, then the scale values of velocity should be linearly related at each age. A curvilinear relation between scale values for an age group would violate scale convergence and would indicate the possibility of a nonlinear judgment function and the need to reevaluate the conclusion regarding age change in the integration function.

The scale convergence criterion also holds some promise for studying developmental changes in social judgments. Birnbaum (1974a) used the scale convergence criterion (as well as the scale-free method) to help eliminate the possibility that a nonlinear judgment function accounted for significant interactions in judgments of likeableness based on combinations of trait adjectives. In Birnbaum's Experiment III, adults judged (a) the likeability of persons described by two adjectives combined and (b) the difference in likeability between pairs of persons each described by a single adjective. The scale convergence test across these two types of judgments supported the interpretation that the interaction could be attributed to the information integration process rather than the judgment function. Analogous experiments seem to be feasible with children for social judgments of many types.

Manipulation of the Judgment Function

One other method with the potential to separate the judgment function from the information integration process is to attempt to manipulate the judgment function. Theoretically, this can be done by manipulating the distribution of stimuli presented (Anderson, 1975; Birnbaum, 1974b, 1982a; Birnbaum et al., 1971; Parducci, 1974; Restle & Greeno, 1970, pp. 157–167). An example of research with adults is provided by Mellers and Birnbaum (1983). In this study subjects judged the overall performance of hypothetical students described by either one or two exam scores. The distributions of exam scores presented varied between subjects but were either positively or negatively skewed. Embedded in the skewed distributions were some stimuli (a factorial design of Exam 1 by Exam 2 values) presented in all conditions of the experiment.

The results for these trials depended on the stimulus distribution in which they appeared. In both conditions significant interactions were obtained, the positively skewed distribution showing an interaction in which the curves converged toward the right, and the negatively skewed distribution showing an interaction in which the curves diverged toward the right. Under the assumption that the judgment functions are linear, these two different interaction patterns would be interpreted as evidence of different information integration functions. Mellers and Birnbaum (1983) however, showed that the data from both conditions were consistent with the interpretation that the information integration function is additive, the scale values are the same across conditions, and the judgment function varies with condition in the way predicted by Parducci's (1974) range–frequency theory. This variation in the judgment function was obtained despite the use of the same end-anchor stimuli in all conditions.

Manipulation of the stimulus distribution is another example of a procedure that usually requires many stimulus combinations. Mellers and Birnbaum (1983), for example, presented 175 stimuli to each subject, a number that seems impractical for research with children younger than approximately age 12. Although smaller designs can be used, it is also desirable to evaluate scale convergence in these experiments, making smaller designs weaker.

Development of response scale use. An interesting aspect of the manipulation of the stimulus distribution is that it may provide a way of studying the development of skills in using rating scales. The judgment function J in Figure 1 should not be relegated to the status of a nuisance variable because it represents the important psychological process of mapping a subjective continuum onto a response continuum. Although this issue is of great importance to the widespread use of rating scales in psychology, with a few exceptions (Anderson, 1975; Attneave, 1962; Birnbaum, 1974b; Helson, 1964; Parducci, 1968, 1974, 1982) it has been quite widely ignored.

The most clearly articulated and supported approach to the problem of the judgment function is Parducci's (1963, 1982) range–frequency theory. Range–frequency theory proposes that rating scale responses are a compromise between an attempt to assign approximately the same number of stimuli to each of the available rating scale categories and the tendency to assign an equal subjective range of stimuli to each category. Very few rating scale studies with children have manipulated the stimulus distribution, in spite of the fact that range–frequency theory provides the most promising framework for studies of how children use rating scales. One exception is an unpublished study by Surber (1978) in which the distribution of stimuli was manipulated in children's judgments of length. The results showed the type of effects predicted by range–frequency theory for kindergarten children and 2nd- and 5th-grade students.

Surber's (1978) results suggest that the same variables may influence both children's and adults' use of rating scales. If range–frequency theory continues to be a viable theory of the judgment function across a wide age range, it will be useful in predicting situations in which age variation in the judgment function will be expected. For example, age changes in the scale values of a variable would be expected to result in some concomitant age changes in the shape of judgment function. There is much further research needed on the variables that may influence the judgment function in children's ratings, however. For example, there may be developmental changes in the psychophysical function for number that would be expected to influence children's magnitude estimations (Attneave, 1962). Also, developmental changes in memory for the stimuli and the response scale categories used would be expected to modify the effects of the stimulus distribution on the judgment function.

Summary

The methods outlined for separating the information integration function from the judgment function are not simple. Implicit in the presentation is the theme that greater leverage is gained in constraining the interpretations of a set of data as the theoretical and empirical network is enriched. The scale convergence criterion and two-operation designs aid in distinguishing between the information integration function and the judgment function. This is because, in part, their experimental designs generate a richer set of empirical relations than that of a simple two-variable factorial design. If the functional measurement approach is to continue its usefulness to developmental psychology, it is important that researchers begin to use designs that allow separation of the components of the judgment process as outlined in Figure 1.

Individual Subject Analyses

Anderson and Butzin (1978) recently reminded developmental researchers that group means may not reflect the individual patterns of data, a fact that has been called to the attention of psychologists many times (Hayes, 1953; Melton, 1936; Sidman, 1952; Skinner, 1959). Thus far, the discussion has focused on the issue of inferring differences among age groups as if each age group showed a homogeneous pattern of response. This will not always be the case, in part because development proceeds at different rates for different individuals. Also, it is possible for individuals to centrate on different stimulus dimensions, with the group data appearing to show a combination of the two dimensions. Alternatively, it is possible for individuals to use a stimulus dimension in opposite ways, with the group data showing no effect of that dimension.

The logic and methods outlined for separating the information integration function from the judgment function can be applied to individuals as well as to age groups. For example, if two individuals give ratings with identical rank orders, an investigator who infers that these individuals differ in the way information is combined would be required to make rather strong assumptions about the judgment function. These points will not be reiterated here. The following discussion considers ways to determine the patterns of the individual data when few replications are available from each subject. All the approaches deserve careful scrutiny in developmental research because it is likely that error variability is related to age.

Analysis of Variance

Analysis of variance can be applied to each subject's data separately, with the results used to draw conclusions about the cues used and the information integration process of each subject. The number of subjects showing significant main effects and interactions at each age can be enumerated, and the age differences can be tested with chi-square or another statistic appropriate for enumerations. When two or more replications of the design have been judged by each individual, replication can be used as the error term. This approach is illustrated by Kun et al.'s (1974) Experiment 3, in which the main effects of ability and effort, as well as the interaction and bilinear component of the interaction were tested. Kun et al. found that the majority of kindergarten children used both ability and effort in judging performance, disconfirming the Piagetian centration hypothesis for this age. Because these analyses were based on only two replications, Kun et al. set the alpha level to .10 in order to increase power. This increase in risk of a Type I error seems preferable in order to decrease the chance of Type II error. However, other studies using individual subject analyses of variance with only two or three replications per subject have adhered to the conventional .05 level (Gupta & Singh, 1981; Wilkening, 1980).

When each individual judges only one replication of the experimental design, it is still possible to carry out an analysis of variance on each subject's data to test the main effect of each manipulated variable. However, the interaction term must then be used as the error term. This is not desirable if interactions are expected, of course.

An obvious drawback of statistical analysis of individual data patterns is that the power of the significance tests will vary with age. A well-known problem in developmental research is that the error terms tend to decrease with increases in age. As Anderson (1980) noted, there appears to be no satisfactory solution to this problem. Consequently, the use of individual subject statistical analyses to test for centration or for use of different types of information will be biased in favor of finding age differences (see the discussion by Anderson & Butzin, 1978). Actually, tests of significance for group data will also be biased in favor of finding that older children use a type of information or use a more complex information integration process, whereas younger children do not. This is a general problem that is often overlooked in developmental research. One partial solution is for investigators to calculate and report the statistical power of the tests at each age level, the effect sizes, or at least the mean squared error.

Centration. A special problem in statistically detecting what Piaget termed *centration* occurs when (a) each subject judges only one replication of an experimental design and (b) it is plausible that over the trials of a factorial design an individual may base a judgment on first one variable and then another, but never both. Anderson and Cuneo (1978a) noted that if this were the case within an age group, the variances of the judgments of stimuli composed of similar valued components should be lower than the variances of the other stimuli. For example, the variances of the judged area of the 7×7 cm and 11×11 cm stimuli should be lower than the variances of the 7×11 cm and 11×7 cm stimuli. This prediction of the centration hypothesis can be tested in the group data using the usual F tests for variances.

Nonstatistical Approaches

Cue usage or integrational capacity. In an attempt to avoid the bias that might be introduced by age differences in the reliability of the data, Anderson and Butzin (1978) categorized each child as using or not using a cue by inspecting the marginal means for that factor. If the means for a child differed by more than one rating scale unit, the child was said to have employed that cue. Using this method, Anderson and Butzin found a developmental increase in the number of cues utilized or "integrational capacity." The results corresponded closely to those obtained by using an analysis of variance on each subject's data. This method and other methods of categorizing data that are considered in this section are also subject to unreliability due to age differences in error variance.

Miller (1982) classified individual subjects as using a variable if the marginal means were ordered in the same way as expected on a priori grounds. This method was used in a study of how interest and noise levels are combined to predict learning. Miller concluded that there was a developmental increase between kindergarten and the 5th grade in the use of noise level in predicting learning. Miller's criterion for when a cue was used was considerably more stringent than Anderson and Butzin's, but it also made assumptions about how the subjects would evaluate information (e.g., high noise level implied less learning than low noise level). In situations in which the scale values for two levels of a cue are close together, Miller's method appears to be subject to Type II error due to age change in response variability.

Qualitatively different cue usage. Another problem occurs when children of a single age level use information in different ways to formulate their judgments. This problem would go unnoticed unless individual data were examined. For example, Surber (1980) found no main effect of the manipulated level of effort on children's judgments of ability. Inspection of the individual data patterns showed that subjects were using effort in two different ways; some judged ability to increase as effort increased, and others judged ability to decrease as effort increased. Because each subject judged only one replication of the design, statistical analyses were not used. Instead, subjects were categorized as showing a positive or negative effect of effort by taking the difference between their judgments of the highest and lowest effort cue combined with each of four levels of performance information. The direction of the majority of the differences was used to categorize each individual. Because a 4×4 (Effort Level $\times$ Performance Level) factorial design was used, the categorization did not involve the two middle levels of effort. These data could be analyzed as a 2×4 (Effort Level $\times$ Performance Level) design to make sure that the categorization did not merely capitalize on chance. The groups differed in the expected way. These directional effects cannot be attributed to group differences in the judgment function because they dramatically alter the rank order characteristics of the data.

There are other examples in which directional differences in use of a cue occur. Butzin (1979) found that some 7-year-olds judged the goodness of a person to increase as function of the person's ulterior motive for a helpful act (the more money a child was offered for helping, the better that child was judged to be), whereas other 7-year-olds judged the goodness of a person to decrease as a function of the ulterior motive. For the 7-year-olds as a group, there was no significant effect of ulterior motive. Another possible example is a study by Wilkening (1981) in which he reported that 5-year-olds showed no significant effect of time on their judgments of speed. Wilkening did not report analyses of individual subject data, and it would be interesting to know whether there are individual differences in the directional effects of the time cue.

Both Wilkening's (1981) and Butzin's (1979) data were obtained from stimuli generated by a 3×3 factorial design. Had they categorized subjects in a way analogous to that of Surber (1980), there would not have been two levels of the variable left for statistical analysis that were independent of the categorization. If these investigators had used more than three stimulus levels, they could have categorized the subjects and then conducted statistical tests on the groups. By using more stimulus levels, researchers would leave open a method for finding and confirming individual differences in information valuation when they do exist.

Another problem of individual differences within an age group was encountered by Leon (1980) in his study of intentionality and consequences in moral judgments. Leon found evidence that children used different strategies for combining intent and consequences. He classified the integration rule of each individual as "configural" (subjects who used the consequence cue only if the intent cue was not described as accidental), intent only, consequence only, or multiplying, based on their responses to a subset of the stimuli. The responses to stimuli not involved in the classification tended to follow the same pattern, supporting the classification criteria. Although Leon could have used statistical analyses to confirm the group differences in strategies, no statistics were reported.

An important issue in Leon's individual difference groups is whether they can be interpreted as reflecting differences in information integration or whether they can be attributed to different judgment functions. Inspection of the rank order characteristics of Leon's groups shows clearly that the differences cannot be attributed solely to the judgment function. A different interpretive problem would have arisen had this not been the case.

Summary

None of the methods for analyzing individual data patterns is completely satisfactory. Nevertheless, developmental data should be routinely examined for individual differences. It is recommended that experimental designs larger than 3×3 factorial be used to facilitate verification of individual differences in the effects of information. When inferring individual differences in the information integration function, the possibility that the judgment function differs between individuals must also be considered.

Conclusions

An examination of theoretical and practical issues in use of mathematical models of judgment in developmental research shows that this approach continues to hold promise. The major advantage of mathematical models of judgment is that they can provide precise theoretical descriptions of either developmental changes or individual differences in the way information is combined, a fundamental issue in the study of cognition. In order to describe successfully either developmental or individual differences in the way information is combined, however, it is necessary to allow the possibility that there are also developmental changes in the way psychological impressions are translated into overt responses. A review of techniques for separating these two aspects of the judgment process shows that although more complex designs are required, the inferential leverage obtained makes the use of such techniques well worth the effort.

References

Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review, 77,* 153–170.

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review, 78,* 171–206.

Anderson, N. H. (1973). Comments on the articles of Hodges and of Schoneman, Cafferty, and Rotton. *Psychological Review, 80,* 88–92.

Anderson, N. H. (1975). On the role of context effects in psychophysical judgment. *Psychological Review, 82,* 462–482.

Anderson, N. H. (1977). Note on functional measurement and data analysis. *Perception & Psychophysics, 21,* 201–215.

Anderson, N. H. (1979). Algebraic rules in psychological measurement. *American Scientist, 67,* 555–563.

Anderson, N. H. (1980). Information integration theory in developmental psychology. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children* (pp. 1–45). Hillsdale, NJ: Erlbaum.

Anderson, N. H., & Butzin, C. A. (1978). Integration theory applied to children's judgments of equity. *Developmental Psychology, 14,* 593–606.

Anderson, N. H., & Cuneo, D. O. (1978a). The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General, 107,* 335–378.

Anderson, N. H., & Cuneo, D. O. (1978b). The height + width rule seems solid: Reply to Bogartz. *Journal of Experimental Psychology: General, 107,* 388–392.

Attneave, F. (1962). Perception and related areas. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 4, pp. 619–659). New York: McGraw-Hill.

Birnbaum, M. H. (1974a). The nonadditivity of personality impressions. *Journal of Experimental Psychology, 102,* 543–561.

Birnbaum, M. H. (1974b). Using contextual effects to derive psychophysical scales. *Perception & Psychophysics, 15,* 89–96.

Birnbaum, M. H. (1978). Differences and ratios in psychological measurement. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory: Vol. 3* (pp. 33–74). Hillsdale, NJ: Erlbaum.

Birnbaum, M. H. (1979). Reply to Eisler: On the subtractive theory of stimulus comparison. *Perception & Psychophysics, 25,* 150–156.

Birnbaum, M. H. (1980). Comparison of two theories of "ratio" and "difference" judgments. *Journal of Experimental Psychology: General, 109,* 304–319.

Birnbaum, M. H. (1982a). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–485). Hillsdale, NJ: Erlbaum.

Birnbaum, M. H. (1982b). On rescaling data to fit the model and concluding that the model fits: A note on monotone transformation. *Perception & Psychophysics, 32,* 293–296.

Birnbaum, M. H. (1982c). Problems with so-called "direct" scaling. In J. T. Kuznicki, R. A. Johnson, & A. F. Rutkiewic (Eds.), *Selected sensory methods: Problems and approaches to hedonics, ASTM STP 773* (pp. 34–48). New York: American Society for Testing and Materials.

Birnbaum, M. H., Parducci, A., & Gifford, R. K. (1971). Contextual effects in information integration. *Journal of Experimental Psychology, 88,* 158–170.

Birnbaum, M. H., & Veit, C. (1974). Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics, 15,* 7–15.

Bogartz, R. S. (1976). On the meaning of statistical interactions. *Journal of Experimental Child Psychology, 22,* 178–183.

Bogartz, R. S. (1978). Comment on Anderson and Cuneo's "The height + width rule in children's judgments of quantity." *Journal of Experimental Psychology: General, 107,* 379–387.

Bogartz, R. S., & Wackwitz, J. H. (1970). Transforming response measures to remove interactions or other sources of variance. *Psychonomic Science, 19,* 87–89.

Busemeyer, J. R. (1980). Importance of measurement theory, error theory, and experimental design for testing the significance of interactions. *Psychological Bulletin, 88,* 237–244.

Butzin, C. (1979). Children's moral judgments of ulterior motives. *Abstracts of Individual Papers, Biennial Convention of the Society for Research in Child Development, 2,* 34.

Butzin, C. A., & Anderson, N. H. (1973). Functional measurement of children's judgments. *Child Development, 44,* 529–537.

Cuneo, D. N. (1980). A general strategy for quantity judgments: The height + width rule. *Child Development, 51,* 299–301.

Cuneo, D. N. (1982). Children's judgments of numerical quantity: A new view of early quantification. *Cognitive Psychology, 14,* 13–44.

DePaulo, B. M., Jordan, A., Irvine, A., & Laser, P. (1982). Age changes in the detection of deception. *Child Development, 53,* 701–709.

Dulany, D. E. (1968). Awareness, rules, and propositional control: A confrontation with S–R behavior theory. In T. Dixon & D. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 340–387). Englewood Cliffs, NJ: Prentice-Hall.

Flavell, J. H., Beach, D., & Chinsky, J. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37,* 283–300.

Grueneich, R. (1982). The development of children's integration rules for making moral judgments. *Child Development, 53,* 887–894.

Gupta, M., & Singh, R. (1981). An integration–theoretical analysis of cultural and developmental differences in attribution of performance. *Developmental Psychology, 17,* 816–825.

Hayes, K. J. (1953). The backward curve: A method for the study of learning. *Psychological Review, 60,* 269–275.

Heller, K. A., & Berndt, T. J. (1981). Developmental changes in the formation and organization of personality attributions. *Child Development, 52,* 683–691.

Helson, H. (1964). *Adaptation-level theory.* New York: Harper & Row.

Hendrick, C., Franz, C. M., & Hoving, K. L. (1975). How do children form impressions of persons? They average. *Memory & Cognition, 3,* 325–328.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* (A. Parsons & S. Milgram, Trans.). New York: Basic Books. (Original work published 1955)

Krantz, D. H. (1974). Measurement theory and qualitative laws in psychophysics. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Measurement, psychophysics and neural information processing* (pp. 160–199). San Francisco: Freeman.

Krantz, D. H., & Tversky, A. (1971). Conjoint-measurement analysis of composition rules in psychology. *Psychological Review, 78,* 151–169.

Kruskal, J., & Carmone, F. (1969). MONANOVA: A FORTRAN-IV program for monotone analysis of variance. *Behavioral Science, 14,* 165–166.

Kun, A. (1977). Development of the magnitude-covariation and compensation schemata in ability and effort attributions of performance. *Child Development, 48,* 862–873.

Kun, A., Parsons, J., & Ruble, D. (1974). Development of integration processes using ability and effort information to predict outcome. *Developmental Psychology, 10,* 721–732.

Leon, M. (1980). Coordination of intent and consequence information in children's moral judgments. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children* (pp. 71–97). Hillsdale, NJ: Erlbaum.

Leon, M. (1982a). Rules in children's moral judgments: Integration of intent, damage, and rationale information. *Developmental Psychology, 18,* 835–842.

Leon, M. (1982b). Extent, multiplying and proportionality rules in children's judgments of area. *Journal of Experimental Child Psychology, 33,* 124–141.

Leon, M. (in press). Rules mothers and sons use to integrate intent and damage information in their moral judgments. *Child Development.*

Marks, L. E. (1974). On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as a science. *Perception & Psychophysics, 16,* 358–376.

Marks, L. E. (1982). Psychophysical measurement: Procedures, tasks, scales. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 43–71). Hillsdale, NJ: Erlbaum.

Mellers, B., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 582–601.

Mellers, B., & Birnbaum, M. H. (1983). Contextual effects in social judgment. *Journal of Experimental Social Psychology, 19,* 157–171.

Melton, A. W. (1936). The end-spurt in memorization curves as an artifact of the averaging of individual curves. *Psychological Monographs, 47*(2, Whole No. 212).

Miller, P. H. (1982). Children's and adults' integration of information about noise and interest levels in their judgments about learning. *Journal of Experimental Child Psychology, 33,* 536–546.

Norman, K. L. (1976). A solution for weights and scale values in functional measurement. *Psychological Review, 83,* 80–84.

Norman, K. L. (1981). Assessing the importance of environmental and societal factors: A way to separate relative importance and subjective value. In W. Molt, H. A. Hartmann, & P. Stringer (Eds.), *Advances in economic psychology* (pp. 37–61). Heidelberg, Germany: Meyn.

Parducci, A. (1963). Range-frequency compromise in judgment. *Psychological Monographs, 77*(2, Whole No. 565).

Parducci, A. (1968). The relativism of absolute judgment. *Scientific American, 219,* 84–90.

Parducci, A. (1974). Context effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception: Vol. II. Psychophysical judgment and measurement* (pp. 127–141). New York: Academic Press.

Parducci, A. (1982). Category ratings: Still more contextual effects. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 89–105). Hillsdale, NJ: Erlbaum.

Piaget, J. (1965). *The moral judgment of the child.* (M. Gabain, Trans.). New York: Free Press.

Restle, F., & Greeno, J. G. (1970). *Introduction to mathematical psychology.* Menlo Park, CA: Addison-Wesley.

Rule, S. J., & Curtis, D. W. (1980). Ordinal properties of subjective ratios and differences: Comment on Veit. *Journal of Experimental Psychology: General, 109,* 296–300.

Schonemann, P. H., Cafferty, T., & Rotton, J. (1973). A note on additive functional measurement. *Psychological Review, 80,* 85–87.

Shaklee, H. (1976). Development of inferences of ability and task difficulty. *Child Development, 47,* 1051–1057.

Shatz, M. (1978). The relationship between cognitive processes and the development of communication skills. In C. B. Keasey (Ed.), *Nebraska symposium on motivation 1977* (Vol. 25, pp. 1–42). Lincoln: University of Nebraska Press.

Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin, 49,* 263–269.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8,* 481–520.

Siegler, R. S. (1978). *Children's thinking: What develops?* Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Richards, D. D. (in press). The development of two concepts. In C. Brainerd (Ed.), *Recent advances in cognitive developmental theory.* New York: Springer-Verlag.

Singer, J. B., & Flavell, J. H. (1981). Development of knowledge about communication: Children's evaluations of explicitly ambiguous messages. *Child Development, 52,* 1211–1215.

Singh, R., Sidana, U., & Saluja, S. (1978). Playgroup attractiveness studied with information integration theory. *Journal of Experimental Child Psychology, 25,* 429–436.

Skinner, B. F. (1959). A case history in scientific method. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2, pp. 359–379). New York: McGraw-Hill.

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology, 54,* 337–341.

Surber, C. F. (1977). Developmental processes in social inference: Averaging of intentions and consequences in moral judgment. *Developmental Psychology, 13,* 654–665.

Surber, C. F. (1978, August). *The use of context effects to examine developmental changes in response processes.* Paper presented at the annual meeting of the American Psychological Association, Toronto, Ontario, Canada.

Surber, C. F. (1980). The development of reversible operations in judgments of ability, effort and performance. *Child Development, 51,* 1018–1029.

Surber, C. F. (1982). Separable effects of motives, consequences, and presentation order on children's moral judgments. *Developmental Psychology, 18,* 257–266.

Torgerson, W. S. (1961). Distances and ratios in psychological scaling. *Acta Psychologica, 19,* 201–205.

Verge, C. G., & Bogartz, R. S. (1978). A functional measurement analysis of the development of dimensional coordination in children. *Journal of Experimental Child Psychology, 25,* 337–353.

Weiss, D. J., & Anderson, N. H. (1972). Use of rank order data in functional measurement. *Psychological Bulletin, 78,* 64–69.

Wellman, H. M., Collins, J., & Glieberman, J. (1981). Understanding the combination of memory variables: Developing conceptions of memory limitations. *Child Development, 52,* 1313–1317.

Wilkening, F. (1979). Combining of stimulus dimensions in children's and adults' judgments of area: An information integration analysis. *Developmental Psychology, 15,* 25–33.

Wilkening, F. (1980). Development of dimensional integration in children's perceptual judgment: Experiments with area, volume, and velocity. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children* (pp. 47–69). Hillsdale, NJ: Erlbaum.

Wilkening, F. (1981). Integrating velocity, time and distance information: A developmental study. *Cognitive Psychology, 13,* 231–247.

Wimmer, H., Wachter, J., & Perner, J. (1982). Cognitive autonomy of the development of moral evaluation of achievement. *Child Development, 53,* 668–676.

Wohlwill, J. F. (1973). *The study of behavioral development.* New York: Academic Press.

Received February 8, 1983
Revision received June 13, 1983 ■