An Exegesis of Two Theories of Compensation Development: Sequential Decision Theory and Information Integration Theory

DENNIS D. KERKMAN AND JOHN C. WRIGHT

Center for Research on the Influences of Television on Children, Department of Human Development, University of Kansas

Previous studies on the development of problem-solving skills in compensation tasks suggest that sequential decision theory more accurately describes the earlier phases of development than information integration theory. Compared to information integration theory, sequential decision theory appears to provide a more valid representation of children's and adults' processing of simpler compensation problems. On more developmentally advanced compensation problems, adults appear to use a variety of strategies. Information integration is capable of describing many of these patterns accurately. However, detailed analyses of performances by children and adults indicate that processing compensation problems by algebraic integration is the exception, rather than the general rule.

INTRODUCTION

In his writings on operational thought (e.g., 1926, 1929, 1970), Piaget stresses the importance of decentration, of considering more than one aspect of a problem at a time. Whether or not one accepts that aspect of Piagetian stage theory, the case has been well made that much of the child's thinking changes, both qualitatively and permanently when she begins to consider the interaction of multiple causes of final outcomes, as is the case in solving compensation problems. The only other changes that have as much portent for cognitive development are the acquisition of language, the development of formal logic, and perhaps the discovery of the permanence of objects and the objectivity of space, time, and causality.

The development of scientific information-processing skills is the focus...
of growing national attention (Carnegie Forum, 1986). Educational policymakers have turned to cognitive psychology for scientific bases to improve the development of problem-solving skills in science and mathematics (National Science Foundation, 1983). The value of developmental models in instructional design is now becoming more fully recognized (e.g., Case, Sandieson, & Dennis, 1986). The recent trend in cognitive theory has been away from grand, unifying theories (e.g., Piaget, 1970) and toward specific models for specific knowledge domains (e.g., Chi, 1978; Glaser, 1982; Glaser & Takanishi, 1986). Increasingly, researchers are confronted with several models of development for the same tasks. For those who focus on basic research in cognitive development as well as for those who seek psychological solutions to these new educational challenges, the choice of models is a crucial decision.

The purpose of models is more than to predict performance. No model will serve to predict performance on all tasks for all children. But the enterprise of generating descriptive models is itself an attempt to find descriptors that transcend task and individual to a degree. When a model cleanly separates types of individuals on the basis of cognitive skills and knowledge as to which ones follow the model and which do not, and when a model can similarly serve to sort tasks into those well analyzed by the model and those not so well analyzed, then that model is useful in describing a domain within cognitive development. When the model also has some correspondence with the processes responsible for development of those skills, then it becomes a potentially useful tool for guiding the design of instruction for those children on those tasks.

Some criteria are identified here that should be considered when choosing among theoretical models of problem-solving development, focusing especially on process validity (Butterfield, 1979) and instructional implications. These criteria are used to compare two leading theoretical models of problem-solving development on multidimensional tasks that involve compensation relationships between offsetting antecedent variables. This comparison also illustrates how the model one chooses can affect the hypotheses one tests, the methods one chooses, and the interpretations one offers for the advancement of cognitive theory and the improvement of science instruction.

The cognitive advance from unidimensional to multidimensional problem-solving is a critical step both developmentally and instructionally. The theories addressed here have been proposed to describe problem-solving on tasks where the outcome is determined by a potentially compensatory relationship between two dimensions of the task. For example, the area of a rectangle depends on both its width and its height. Similarly, the mechanical advantage of a lever depends on both the force exerted and its distance from the fulcrum. Two major theoretical explanations for
children's development of the ability to solve such problems are compared here, and recent research concerning their validities is reviewed.

At a very general level, the two theories represent fundamentally different schools of thought concerning the basic nature of developmental progress. "Sequential decision theory" (SDT) (Siegler, 1976) is one of several information-processing accounts of problem-solving development that emerged in the late 1970s and early 1980s within the general framework of "rule sampling theory" (e.g., Brainerd, 1979, 1981; Wilkenson, 1982a, b). The rule sampling approach holds that the highly consistent behavior patterns often observed on Piagetian problem-solving tasks are the result of relatively simple rules that are stored in the child's long-term memory and sampled from memory to solve the problem at hand. In this view, two major sources of developmental change are identified: changes in the retrieval probabilities of the rules, and changes in the sophistication of the rules themselves. SDT focuses mainly on the latter.

Sequential decision theory maintains that children progress through a series of discrete and distinct knowledge states that culminate in a full scientific level of knowledge about a task. The fully ramified rule sampling theory (e.g., Brainerd, 1981) is stochastic, thus providing a view of developmental change that can be continuous over relatively long periods. Nevertheless, at the level of processing of an individual problem, SDT describes sequential, discrete steps in the processing of distinct types of problems, as opposed to a single process that combines stimulus values in the same way on all problems.

Unlike SDT, Anderson's (1981, 1982) "information integration theory" (IIT) is derived from a perspective that finds its historical roots in psychophysics and learning theories (e.g., Coombs, Dawes, & Tversky, 1970; Stevens, 1951). Anderson (1981, 1982) has provided summaries of this research.

Information integration theory assumes continuity not only of developmental change, but also at the level of the processing of an individual problem; the transformation of the stimulus configuration into an overt response is represented as a single, continuous function for integrating stimulus values into responses for all problems. (Anderson, 1981; Cuneo, 1982). Also unlike SDT, IIT does not imply a taxonomy of "problem types."

The purpose here is not so much to choose between theories or differentially evaluate as it is to delineate the conditions of task type, levels of cognitive development, and purposes for which each model works best. From the view of rule sampling theory, the growth of models for two-dimensional problem-solving presents the possibility for showing how a variety of task specific minitheories can be integrated in such a way that they can be applied not only to many tasks and children, but to many
different response formats such as continuous adjustment, estimation, prediction, or categorical judgment. The models compared here are prototypic for any kind of multidimensional problem-solving. The fundamental differences between bottom-up and top-down processing, between empirical, experiential-based trial and error, and deductions of logical necessity are well exemplified in these contrasting answers to that question first raised by Piaget: How does a child go about working with more than one kind of information at a time?

THE TASKS

Each of the theories represents a formal attempt to provide a unified explanation of how children develop the ability to solve problems that involve compensatory relationships. It is important at the outset to distinguish the task, or problem-to-be-solved, from the various strategies that individuals may apply to solve the problem at various times in their development. Certain features of the problem situation per se impose constraints on the kinds of actions that will be more or less successful but there is generally more than one way to achieve the goal, depending mainly on the problem-solver's strategies.

The basic structure of each task involves offsetting relations between two antecedents and an outcome. To maintain a given outcome \( \text{[C]} \) unchanged, increases in antecedent \( \text{[A]} \) can be counteracted by decreases in antecedent \( \text{[B]} \). Otherwise, a change in \( \text{[C]} \) is caused by any change in \( \text{[A]} \) or \( \text{[B]} \) or both. Whether a given task has the logical structure of a compensation task is a matter of physical and/or logical fact. But whether a task is perceived as being of this kind is determined by the perceptions, knowledge base, and cognitive skills of the problem-solver.

The tasks addressed by these theories involve differences, rates, ratios, proportions, or other multidimensional interactions. In addition to the examples just mentioned concerning rectangles and levers, consider these tasks: the distance travelled by an object depends on both its speed and the time elapsed; the size of the shadow that an object casts against a wall depends on both the size of the object and its distance from the light source and the reflecting surface; the probability of drawing an ace depends on both the number of aces and the total number of cards remaining in the deck. (For recent studies of compensation in social attributions, see Surber, 1984a, b, c; 1985a, b; Surber & Haines, in press.)

Piaget and his collaborators (Inhelder & Piaget, 1958; Piaget, Inhelder, & Szeminska, 1960) originally described the development of the ability to solve compensation problems, and in doing so distinguished tasks involving additive compensation (e.g., length and number conservation) from those involving multiplicative compensation (e.g., area and volume con-
For example, if two lines are initially the same length, and the length of one is increased, the change is additive because it yields a unit-for-unit increase in the difference between the lines.

Such additive problems are distinguished from multiplicative problems wherein the antecedent variables interact to determine the outcome. If two rectangles are initially identical, and the height of one is increased, then the magnitude of the effect on area is not unit-for-unit, but depends on its width. When one is told that the outcome (area) is to be constant, the job is to reason that when there is an increase in height there exists a reduction in width that provides an opposing offset, leaving area unchanged. The logical certainty, without actually trying it, that there exists an exactly compensating offset that could be achieved by changing the other antecedent in a specified direction is an early formal operational skill. Comprehending the simultaneous operation of opposing actions is what the Genevans call qualitative compensation, as distinct from quantitative compensation, the latter involving "an actual proportion to which numbers can be assigned" (Inhelder & Piaget, 1958, p. 219).

Clearly, the logical certainty of the possibility of constructed compensation is a more advanced cognitive achievement than the logical appreciation of invariance involved in conservation. Several investigators have found that compensation training is sufficient for the appearance of conservation (see Silverman & Rose, 1982). But it is equally clear that a concrete form of compensatory logic ought to follow the development of awareness of more than one antecedent variable and precede the fully abstract form of compensation logic. Thus we shall not be surprised to find evidence of (1) univariate logic in young children, followed in order by (2) discovery of a second causal variable, (3) appreciation of their simultaneous and potentially opposite effects, and somewhat later (4) the realization of the logical necessity for the existence of an exactly compensating offset in one antecedent for any imposed change in the other.

**STUDIES OF STRATEGY DEVELOPMENT ON COMPENSATION TASKS**

SDT was proposed by Siegler (1976), based partially on his critical reanalysis of Inhelder and Piaget's (1958) classic studies of *The growth of logical thinking from childhood to adolescence*. SDT was first used to describe developmental changes in children’s knowledge of the proportional compensation between weight and distance from the fulcrum in determining which end of a balance scale will tip down. SDT was then used to describe children’s development of the ability to judge the relative fullness of vessels (cf. Bruner & Kenney, 1966; Siegler & Vago, 1978) and time, distance, and speed concepts (cf. Piaget, 1971; Siegler & Richards, 1979).
SDT has been used to describe children’s development on covariation problems (Shacklee & Mims, 1981) and problems involving falling objects on inclined planes (cf. Inhelder & Piaget, 1958; Ferretti, Butterfield, Cahn, & Kerkman, 1985). Siegler (1981) used SDT to analyze children’s development across several developmentally sensitive problem-solving tasks: the balance scale, projection of shadows, probability, and conservation of number, liquid quantity, and solid quantity.


Based on these findings, Anderson and Cuneo (1978) and Cuneo (1980, 1982, 1983) have proposed that even preschool-aged children’s judgments of quantities are regulated by a general-purpose additive integration rule, rather than being “centered” on a single stimulus dimension, as has been asserted by the Genevans (e.g., Inhelder & Piaget, 1958) and Siegler (1976, 1981).

Siegler (1981, pp. 100–102) provides a table listing several two-dimensional compensation tasks that have been used in research on SDT. Table 1 is an extended revision of Siegler’s (1981) table; it includes other tasks that have been analyzed using either SDT or IIT to study the development of the ability to solve compensation problems.

Table 1 also lists the solution dimensions for each task. In Table 1, the dimensions that typically have been reported to “dominate” the judgments of young children on these tasks are listed. Consider the additive task of comparing the lengths of two sticks (“A” vs. “B”). Children who are less than 5 or 6 years old typically base their judgments mainly on the ends that are nearer to them. They fail to recognize that a length difference nearer to them can be exactly offset by a difference in the opposite direction at the end farther from them. Such judgments are said to be “centered on” (Inhelder & Piaget, 1958) or “dominated by” (Siegler, 1976) the end nearer the child.

In the remainder of this paper, a multiplicative task called the balance scale will serve as the main example to explain and contrast the two theories. Most of the research that directly concerns the comparison of
TABLE 1
SOME COMPENSATION TASKS IN SIMPLE PHYSICS THAT HAVE BEEN STUDIED USING
SEQUENTIAL DECISION THEORY AND/OR INFORMATION INTEGRATION THEORY

<table>
<thead>
<tr>
<th>Task name</th>
<th>Outcome quantity</th>
<th>Dominant dimension</th>
<th>Subordinate dimension</th>
<th>Correct formula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Additive tasks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>Length ($L$)</td>
<td>Near lag ($N$)</td>
<td>Far lag ($F$)</td>
<td>$L = N + F$</td>
</tr>
<tr>
<td>Time</td>
<td>Time ($T$)</td>
<td>End lag ($E$)</td>
<td>Begin Lag ($B$)</td>
<td>$T = E - B$</td>
</tr>
<tr>
<td><strong>Multiplicative tasks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectangles</td>
<td>Area ($A$)</td>
<td>Height ($H$)</td>
<td>Width ($W$)</td>
<td>$A = H \times W$</td>
</tr>
<tr>
<td>Number</td>
<td>Number ($N$)</td>
<td>Length ($L$)</td>
<td>Density ($D$)</td>
<td>$N = L \times D$</td>
</tr>
<tr>
<td>Liquid</td>
<td>Volume ($V$)</td>
<td>Height ($H$)</td>
<td>Area ($A$)</td>
<td>$V = H \times A$</td>
</tr>
<tr>
<td>Solid</td>
<td>Volume ($V$)</td>
<td>Length ($L$)</td>
<td>Area ($A$)</td>
<td>$V = L \times A$</td>
</tr>
<tr>
<td>Shadows</td>
<td>Size ($S$)</td>
<td>Object ($O$)</td>
<td>Distance ($D$)</td>
<td>$S = O \times D$</td>
</tr>
<tr>
<td>Triangles</td>
<td>Area ($A$)</td>
<td>Height ($H$)</td>
<td>Width ($W$)</td>
<td>$A = \frac{H}{2} \times W$</td>
</tr>
<tr>
<td>Speed</td>
<td>Speed ($S$)</td>
<td>Time ($T$)</td>
<td>Distance ($D$)</td>
<td>$S = D/T$</td>
</tr>
<tr>
<td>Fullness</td>
<td>Fullness ($F$)</td>
<td>Liq. Ht. ($L$)</td>
<td>Glass Ht. ($G$)</td>
<td>$F = L/G$</td>
</tr>
<tr>
<td>Balance scale</td>
<td>Torque ($T$)</td>
<td>Weights ($W$)</td>
<td>Distance ($D$)</td>
<td>$T = W \times D$</td>
</tr>
<tr>
<td>Inclined plane</td>
<td>Velocity ($V$)</td>
<td>Angle ($A$)</td>
<td>Distance ($D$)</td>
<td>$V = \sin(A) \times D$</td>
</tr>
<tr>
<td>Probability</td>
<td>Probability</td>
<td>Frequency</td>
<td>Frequency</td>
<td>$P = D/(D + U)$</td>
</tr>
</tbody>
</table>

Theories has used the balance scale task. The balance scale consists of a crossbar attached at its center to a pedestal so that the crossbar can tip down on either side. The crossbar has six pegs on each arm located at equal distances from each other. Weights are stacked on pegs at various distances from the center. All the weights are identical. In the "prediction" version of this task, children are shown weights on the crossbar with its ends supported by wooden blocks, and asked to predict if the scale will tip down on the left, or the right, or if it will stay balanced when the blocks are removed (Inhelder & Piaget, 1958; Siegler, 1976, 1981). In the "adjustment" version of the balance scale task (Inhelder & Piaget, 1958; Wilkening & Anderson, 1982), a fixed number of weights is placed on the left arm of the scale on a peg near the center, then the child is given a specified number of weights for the right side and asked to adjust their distance to make the scale balance.

LOGICAL AND METHODOLOGICAL COMPARISON OF THE MODELS

The comparison begins with a consideration of the logical potential of the model associated with each theory—its internal consistencies and operating characteristics as a formal system. After assessing the models' logical potentials, the validities of the theories are evaluated, based on available empirical evidence.

The criteria used to compare the SDT and the IIT models are (1) their task domains and basic identifying and distinguishing features, (2) potential descriptive accuracy and falsifiability, (3) potential for being tested by
a variety of independent methods, and (4) potential for describing and predicting the development of these problem-solving skills.

Sequential decision theory and information integration theory both provide clearly communicable formal systems, or "models," for representing the problem-solving processes of individuals. Both theories base their representations of an individual's knowledge not only on the kinds of problems solved correctly, but also on specific patterns of errors committed on specific problems. However, they present fundamentally different conceptualizations of how children process compensation problems and of how these problem-solving processes develop.

IDENTIFYING AND DISTINGUISHING FEATURES

Sequential Decision Model

Siegler (1976) proposed the decision model to describe the development of problem-solving on the balance scale. The model specifies a sequential elaboration of four "rules." At the level of rule 1 (see Fig. 1), decisions are based solely on one "dominant" dimension of the task. On the balance scale, the dominant dimension is usually the number of weights on each side. Rule 1 (weight dominant) describes children who predict that the arm with more weight tips down and that equal amounts of weight always balance—regardless of their distances from the fulcrum.

At rule 2 (see Fig. 2) one branch of the rule-1-level decision structure is elaborated: if the values of the dominant dimension are equal, then the values of the other "subordinate dimension" are tested. On the balance scale, weight differences are assessed first; if there are equal amounts of weight on both arms, the arm with its weights farther from the middle is predicted to tip down.

At rule 3 (see Fig. 3) the other branch of the decision-tree is elaborated; on the balance scale, distances are now compared whether the weights are equal or not. However, when one arm of the crossbar has more weights, but the weights on the other arm are farther from the center, the two dimensions yield conflicting predictions. Rule 3 does not contain a mechanism for accurately compensating between the two dimensions on these "conflict problems." Instead, rule 3 describes a pattern of random guessing on conflict problems.
At rule 4 (see Fig. 4) conflict problems on the balance scale are solved as follows: for each peg occupied by weights, the number of weights is multiplied by their distance from the fulcrum. This product is the invisible, theoretical force that physicists call “torque.” If more than one peg on a side has weights, the total torque for that side is computed by adding the products for each of its pegs. The arm with the greater total torque always tips down, and equal torques always balance.

**Information Integration Model**

The information integration model maintains that on two-dimensional problems both dimensions are taken into account, or “integrated,” and that these integration rules can be modeled by simple algebraic formulas (Anderson, 1981). For example, a child’s pattern might be accurately described by adding the number of weights plus their distance, then com-
paring the sums for each arm, and predicting that the arm with the greater sum will tip down. Some subjects give greater importance to one dimension than the other. The algebraic model explicitly represents these subjective differences as weighting factors in the formula derived for each individual. For example, on the balance scale an individual might subjectively evaluate the weight variable as being three times more important than the distance variable. Algebraically:

\[ V = (3W_L + 1D_L) - (3W_R + 1D_R), \]  

where \( W \) = number of weights; \( D \) = distance; \( L \) = left arm of the balance scale; \( R \) = right arm of the scale; and \( V \) = value. If the value, \( V \), is positive, the scale is predicted to tip down on the left. If \( V \) is negative the scale is predicted to tip down on the right. If \( V \) is zero the scale is predicted to balance.

Multiplicative as well as additive integration rules can be modeled algebraically. For example,

\[ V = (3W_L \times 1D_L) - (3W_R \times 1D_R). \]  

A least-squares analysis of each subject's data is used to assess that subject's weighting values for the two dimensions and to determine em-

\^1 Formulas (1), (2), and (3) differ from those shown in Table 1 only in that they represent the complete computation for both of the objects-to-be-compared as well as the comparison (represented by subtraction), whereas Table 1 shows only the dimensions and the relationship between them for one of the objects-to-be-compared.
pirically whether the subject’s data are better described by an additive or multiplicative function. (Details of the “functional measurement method” that are associated with IIT are discussed under Methods and Precision, below.)

Several basic features distinguish the two theoretical models. SDT identifies a taxonomy of problem types. At the level of the individual problem, processing proceeds according to a sequence of discrete decisions. These permit efficient, line-of-least-effort processing: the least complex problems are diagnosed and solved at the simplest level, while the more complex problems are passed along to more advanced levels until a level of processing appropriate to the problem is encountered or the upper limit of the child’s processing capability is reached. SDT also allows for “fall-back” rules (Siegler, 1983). For example, in the absence of complete information, or under conditions involving high degrees of response uncertainty, a developmentally prior level of processing might be applied.

In contrast, IIT identifies a single algebraic rule for each subject which is a summary description of that subject’s responses on all problems. In this view, problems differ not in kind, but rather in the numerical values they provide as input for the rule that functionally integrates them into responses.

POTENTIAL DESCRIPTIVE PRECISION

Models and Precision

An important criterion for choosing between two models is their potential for describing the domain of behaviors in question. On a typical 24-item prediction test with three response options (left, right, and balance), the number of different response patterns is $3^{24}$ (more than 2.82 billion). On a typical balance scale adjustment task with six positions and a maximum of four weights per position, there are $6^{24}$ possible response patterns. The algebraic model’s mathematical flexibility means that a number of these patterns can be described with a high degree of accuracy using integration rules. For example, assume for the sake of argument that the weighting factors for the two dimensions were restricted to the integers from 1 to 5 (in fact, they could take any real number values). Assuming integers only, there would be 19 distinguishable weighting

\footnote{In a footnote, Wilkening and Anderson (1982) acknowledge that nonintegration rules may be applied on problems where only one dimension differs. They restrict their discussion to “conflict” problems, and then proceed to criticize the use of “decision-tree methodology” on the grounds that the rules diagnosed depend on the problems used.}
334  KERKMAN AND WRIGHT

combinations. For each of these, the rule that integrates them might be additive or multiplicative, for a total of 38 "ideal" linear rules.

The sequential model is comparatively impoverished in this regard. Assuming that either dimension could be dominant, the sequential model specifies a total of six rules that can be identified with perfect accuracy (rules 1 and 2 could have either dimension dominant), whereas the dominant dimension cannot be distinguished for rules 3 and 4 on the basis of predictions. The sequential model generates comparatively fewer "ideal patterns" than the algebraic integration model and thus, on a priori grounds, it has less potential for describing the variety of behavior patterns that could occur. Later, we shall evaluate how well it describes those patterns that do, in fact, occur.

Methods and Precision

The discriminative precision of the methods used to assess these problem-solving rules has been an issue of much debate. Theoretically, a child who uses sequential rule 1 (weight dominant) should get all of the simple balance problems and all of the simple weight problems correct. Furthermore, the rule 1 child should commit specific types of errors: on problems where both arms have the same amounts of weight, but the weights on one side are farther from the middle (called "simple distance problems"), rule 1 children predict that the scale will balance. On problems where the weight and the distance dimensions lead to opposite predictions ("conflict problems"), rule 1 children should always choose the side with more weights. Rule 2 should produce the same pattern of correct answers and weight-dominated errors, except on simple distance problems, which rule 2 children solve correctly. Rule 3 should produce the same pattern as rule 2, except that both dimensions are considered on conflict problems. However, rule 3 children lack the correct torque rule for integrating the two dimensions and so are forced to "muddle through" (i.e., guess) on conflict problems, thus producing a pattern of errors on conflict problems that is no longer dominated by the weight dimension, but is essentially random. Rule 4 should produce perfect performance on all problems.

Siegler (1976) arbitrarily established a set of criteria for classifying children as using one of these four rules based on the number of problems in each of the six problem types that were answered correctly and on the types of errors that were committed on each type of problem. To be classified as using one of these rules, the child's pattern must conform to the ideal pattern for that rule on 20/24 items (four items × six problem types).

3 All combinations of weighting values that have the same ratio (e.g., 2W + 1D; 4W + 2D) yield identical patterns on prediction or adjustment tasks.
A particular weakness of the sequential model is the chance-level performance predicted for rule 3 on conflict problems. A wide variety of diverse behavior patterns are likely to be obscured within this imprecise catch-all category of subjects whose performance on nonconflict problems is correct, but their pattern of answers on conflict problems is not clearly dominated by either stimulus dimension. It is unlikely that these patterns are purely random guesses (Klahr & Siegler, 1978). The algebraic model’s determinate predictions for conflict problems offer potentially more precision for describing such performances than the stochastic predictions of SDT rule 3.

Wilkening and Anderson (1982) criticized Siegler’s (1976) rule assessment method on the grounds that it is incapable of distinguishing sequential rules from integration rules. For example, suppose the subject used an additive integration rule in which the distance factor was given less subjective importance than the weight factor. Using this rule, distance would appear to influence the subject’s predictions only on those problems where the amounts of weight on each arm are equal and therefore cancel each other out. The resulting pattern of behavior would be misclassified either as sequential rule 2 or rule 3 using Siegler’s (1976) criteria, depending on (a) how much subjective importance is assigned to the two dimensions, and (b) the specific problems used on the test (Wilkening & Anderson, 1982). If the weight dimension is assigned a subjective importance three or four times larger than that of the distance factor, then the additive integration rule generally will be misclassified as a sequential rule 2 because (a) all nonconflict problems will be answered correctly, and (b) the errors on the conflict problems will show a clear-cut bias toward the side with more weights. Such a rule could only produce a response that favors the side with more distance when a very extreme distance on one side is pitted against a very small advantage in weights on the other side.

A response pattern that would be misclassified as sequential rule 3 would be obtained when (a) the subjective weightings of the two dimensions are nearly equal, and (b) the test includes a large proportion of items that involve extreme discrepancies between the two dimensions, so that the pattern of responses on conflict problems does not appear to be dominated by either.

Siegler’s “rule 4” includes a multiplicative algebraic rule for integrating the two dimensions on conflict problems. At this level, SDT and IIT cannot be distinguished on the basis of answer patterns, adjustments, or ratings. Both yield correct solutions, and (for conflict problems at least) they are the same rule.

The information integration theory employs a method for assessing subjects’ rules that is derived from classical statistics. In these studies,
the experimenter systematically manipulates dimensions of the task, and the dependent variable is either (a) the subject's rating of the magnitude of the outcome or (b) the degree of the subject's adjustment of one dimension of the task in order to counteract the experimenter's changes in another dimension. For example, in the adjustment version of the balance scale, the child is given a fixed amount of weight, and then asked to adjust its distance from the fulcrum in order to exactly counterbalance each of the weight $\times$ distance combinations presented on the experimenter's side of the fulcrum (Wilkening & Anderson, 1982). In order to permit single-subject analyses that include the interaction term, each weight $\times$ distance combination is presented at least twice. When these data are subjected to a repeated measures analysis of variance, an additive integration rule is indicated by main effects for both number of weights on the experimenter's side and distance of the experimenter's weights from the fulcrum. However, if the child is focusing on only one of the two dimensions, then there will be only one main effect, and the graphic representation will show the functions for all values of the other dimension as essentially colinear and indistinguishable, as in Fig. 5. If the child is integrating the two dimensions additively, then the child's adjustments to one weight at various distances should run parallel to his/her adjustments to two weights at each of those distances, and so forth, as in Fig. 6. However, if the child is integrating the two dimensions multiplicatively, then the adjustments (or ratings) will show increasing increments as the number of weights to be counterbalanced increases, taking the fan-shaped form shown in Fig. 7. In the terms of analysis of variance, the multiplicative rule is indicated by the linear interaction of weight $\times$ distance because the magnitude of the effect for the weights factor is not constant across levels of the distance factor. If the plot identifies the correct answers for problems, then it is reasonable to say that the subject is using true metric proportionality to solve the problems. If the plot is a linear fan, but its points are not correct answers, then the subject is apparently combining the dimensions by qualitative proportionality (Surber & Haines, in press).

![Fig. 5. Information integration graph of sequential rule 1.](image)
Inhelder and Piaget (1958) viewed qualitative proportionality as penultimate to the development of metric proportionality.

Knowledgeable statisticians (e.g., Cronbach, 1951) have advised against exclusive reliance on the significance of the $F$ ratio to assess individual consistencies, such as rule governedness. It is possible to obtain significant $F$ ratios for weight and distance and yet account for considerably less of the behavior pattern than the 75% concordance criterion that is required by Siegler’s method of rule assessment. Given the degrees of freedom implied by a factorial, repeated-measures design, many of the billions of potential patterns would yield “significant differences” and therefore be classified as integration rules. Thus, algebraic integration is vulnerable to Type I errors—finding a “rule” where none exists. The amount of error tolerated in such statistically defined “rules” makes their validity somewhat dubious unless they are followed up with tests that indicate the actual precision of the algebraic rule in accounting for the individual’s behavior, not just its statistical significance (e.g., Anderson & Cuneo, 1978; Cuneo, 1980). Nevertheless, the algebraic model is capable of describing more of the potential patterns than the sequential model, the statistical method associated with IIT is well-known, and tech-
niques are available for calculating the precision of the fit between the rules and the data.

Popper (1959) argued convincingly that models should be judged relative to their "falsifiability"—how many potential empirical outcomes would contradict the theory? While the descriptive flexibility of IIT means it can yield precise descriptions of many behavior patterns as functions of stimulus dimensions, it also means that IIT is less falsifiable than SDT.

Some outcomes would disconfirm the predictions of IIT. Integration implies that both stimulus dimensions are combined. Therefore, all subjects should have significant main effects for both stimulus dimensions. A zero weighting for one of the dimensions indicates that zero integration is occurring between the dimensions. Zero weightings can occur in some subjects' rules, but properly speaking, these are no longer integration rules because exactly zero integration is occurring in such rules. Thus, one positive attribute of the method associated with IIT is that it can describe conditions that could falsify it.

Sequential decision rule 1 specifies that only the dominant dimension is taken into account. Describing the SDT rule 1 behavior by Anderson's (1981) functional integration method should yield a zero weighting for the subordinate dimension. Rule 2 specifies that the subordinate dimension should be taken into account only when the values of the dominant dimension are equal. Describing rule 2 algebraically should yield a zero weighting for the subordinate dimension on all problems except those where the values of the dominant dimension are equal. Such findings would falsify IIT as a theory for describing this level of problem solving on the task. In contrast, the specificity of the response patterns predicted by SDT and the small number of ideal patterns predicted mean that it is more falsifiable than IIT, especially when predictions about developmental sequence are considered. IIT allows a single developmental transition from additive to multiplicative terms. SDT predicts three transitions and specific order of their occurrence.

INDEPENDENCE OF MODEL AND METHODS

There has been some confusion as to whether Wilkening and Anderson (1982) are critical of the sequential decision model or the method for rule-type assessment. They contended that "... [Siegler's] conclusion that the underlying process is a binary decision tree may be an artifact of the methodology [sic]" (Wilkening & Anderson, 1982, p. 219). Models sometimes become associated with a particular method, but one criterion for judging a model is that it should be testable by a variety of methods (Campbell & Fiske, 1967). Wilkening and Anderson (1982) describe the
THEORIES OF PROBLEM-SOLVING DEVELOPMENT

identification of sequential rules by the functional measurement method of IIT. Ferretti et al. (1985) have demonstrated conclusively that both algebraic and sequential rules can be identified and distinguished from each other using the prediction version of the task and the rule-assessment method outlined by Siegler (1976). Wilkening and Anderson’s (1982) assertions regarding the diagnostic insensitivity of Siegler’s “decision-tree methodology” are not correct. However, it is correct to say that neither method will reveal the other type of rule if the experimenters do not test for it.

A central problem of psychophysics is obtaining equal interval scales of subjective experiences. Anderson states that “Functional measurement, based on the guiding principle of using the model itself as the base and frame for measurement provided a resolution to this difficulty” (Anderson, 1981, p. 3). Far from resolving the difficulty, it would appear that this “reversal of the traditional approach” confounds the model with the measures intended to test it. If one must first assume that stimulus dimensions are integrated according to algebraic formulas in order to obtain any measurements of this purported integration phenomenon, then how can such a theory be independently validated? What external criterion can serve to test it? Anderson (1981) states that the feasibility of algebraic approach depends on many-variable analysis so that there are sufficient degrees of freedom to test the data for goodness of fit to algebraic functions. However, Krantz and Tversky (1971) showed that failures of goodness of fit tests can result from noisy measurement as well as from invalid rules. They describe methods for testing the critical axioms of the structure.

Interpretation of IIT analyses is critically dependent on the nature of the judgment function, which relates the integrated subjective impression to the overt response (Surber, 1984a; Surber & Haines, in press). For example, if the relationship between the additively integrated subjective impression and the overt response were exponential, rather than linear, then additive integration could produce a diverging fan-shaped pattern that is typically considered the hallmark of multiplicative integration, instead of the parallel pattern that is the standard diagnostic indicator of additive integration rules. The fundamental indicator of compensation is the inverse rank ordering of ratings (or adjustments) in one variable in order to offset changes in the other (Surber & Haines, in press). Thus, so long as the judgment function is assumed to be monotonic, both qualitative compensation and metric proportional compensation can be diagnosed by information integration methods.

Surber & Haines (in press) review two procedures for unconfounding the integration and the judgment functions. The methods generally require even more trials per child than the factorial designs necessary for
IIT analyses that assume the judgment function to be linear. For example, Birnbaum and Viet (1974) asked subjects to rate (a) the difference between two weights and (b) the ratio of the two weights (e.g., "half as heavy"). As anticipated, parallel plots were found for the difference task and fan-shaped plots for the ratio task. However, closer analysis of the scale values and a monotonic transformation led to the conclusion that the underlying integration function was actually additive on both tasks, but the judgment function was nonlinear for the ratio judgment task. Surber and Haines (in press) note that the number of trials required is a potential limitation of IIT in the study of children's development. However, the potential of these designs to engender "trials effects" such as warm-up, fatigue, or learning effects has not received sufficient study.

In sum, the algebraic model has not achieved a satisfactory degree of independence between the definitions of its model and the methods of measurement. Alternative methods that provide the possibility for converging evidence from independent types of measurement (e.g., Belmont, Ferretti, & Mitchell, 1982) and analysis of the axiomatic qualitative structure of the data (e.g., Krantz & Tversky, 1971) would greatly improve the utility of the model and fortify the theory.

DESCRIPTING AND PREDICTING DEVELOPMENT

While Siegler set out to produce a developmental description of successive performance levels, Anderson and his colleagues, at least initially, did not. Anderson's work has been extended from adults to children (cf. Anderson & Cuneo, 1978). Wilkening and Anderson (1982) have argued that IIT is superior to SDT for characterizing the rule-governed behaviors of individuals on several developmentally sensitive tasks. Accordingly, IIT should be evaluated as a developmental theory.

The algebraic model does not prespecify any developmental change in the integration function. Instead, changes in the weighting factors or the operations are viewed as empirical questions. Anderson (1981) and his colleagues (Cuneo, 1980, 1982; Wilkening, 1981) have presented evidence suggesting that preschool-aged children use a general-purpose adding rule, even on tasks where a multiplicative rule is correct. However, IIT provides no a priori theoretical rationale for the developmental priority of additive operations. The authors offered two mechanisms to account for these data: one involves averaging the visual estimates for height and width, and the other is a perimeter rule in which the child tracks the perimeter of the figure and uses the total perimeter to determine the response.

Anderson and Cuneo (1978) note that the second mechanism fits the additive data as well as the first, but is in fact a unidimensional, nonin-
integration process. They also note that the fan-shaped pattern usually associated with multiplicative compensation could be produced by an additive process: “the figure could be scanned with a unit square in the mind’s eye and a count made of the number of times the square covered the figure” (Anderson & Cuneo, 1978, p. 363).

Anderson and his colleagues (e.g., Anderson & Cuneo, 1978; Wilkening & Anderson, 1982) have not provided any theoretical mechanism for transition from one algebraic rule to another. In a later study, Cuneo (1982) proposed a single algebraic model for describing developmental changes on these tasks. The model includes both a weighted additive and a weighted multiplicative component for each side of the scale. Algebraically,

\[ V = [(W_L + D_L) + (W_L \times D_L)] - [(W_R + D_R) + (W_R \times D_R)]. \]  

Cuneo (1982) notes that this model is more general than the original (Anderson & Cuneo, 1978) formulation and provides a way to conceptualize the course of development as a gradual and continuous process in which the weighting factors for the multiplicative components gradually increase and those of the additive component gradually decrease. The model allows for examination of age-related differences in the degree of additive and multiplicative components of children’s judgments. She also acknowledged that this model is neutral with respect to the mechanism responsible for the formation of the number percept.

Hence, IIT does not provide a “blueprint” for predicting development on these tasks, it is ambiguous with respect to the integrative or nonintegrative nature of the processes that underlie these purportedly algebraic integration rules, and it does not specify any mechanism of transition between developmentally distinct types of rules. Evoking such a formulation as an explanation for the developmental shift is preformist in the Platonic sense: it presumes that both additive and multiplicative components are already available (with different weightings) from the outset of development. As an explanation for the developmental shift, this reformulation merely begs the question. However, it does permit precise description of a variety of rules that individuals could adopt along their way to acquiring the normative rule for a given task, and thus represents a potentially useful analytic method.

**SUMMARY OF THE COMPARISON OF MODELS**

The sequential model identifies qualitative differences between types of problems and the ways they are processed, whereas IIT focuses exclusively on quantitative differences between problems. The algebraic model is more flexible mathematically, generating a larger number of potential
ideal rules than the sequential model, and the method associated with IIT readily yields a metric for the precision of the fit between data and ideal.

The sequential decision model and the associated prediction method of measurement have been criticized for diagnostic insensitivity to integration rules, and in fact the final level of the decision-tree model is an algebraic integration rule. It is possible to detect integration rules with the methods used by SDT and it is possible to detect sequential rules with the methods used by IIT. However, neither model can detect the "other kind" of rules unless the experimenter looks for them. In other words, the insensitivity is not in the method, it is in the researcher's selection of problems. The algebraic model provides a means for describing changes in individuals, but describing by an algebraic integration rule does not imply that the underlying process is either algebraic, or integrative.

The algebraic model provides no a priori basis for predicting or explaining developmental transitions from one rule to the next. Unless one assumes that the judgment function is linear, IIT requires rather complex experimental designs to disambiguate additive and multiplicative rule classifications.

Surber and Haines (in press) correctly note that the SDT rule assessment approach "...is not well-adapted to the description of continuous or quantitative changes" (p. 52). Within the more general framework of rule sampling theory (e.g., Brainerd, 1981), changes in the probabilities associated with the selection of the rules permit a continuous developmental shift from one rule to the next. IIT is more rich in the number of potential rules it can describe, but its very mathematical flexibility means that it is also potentially less falsifiable than SDT. The latter is also more falsifiable because it predicts a specific sequence of three developmental transitions.

EMPIRICAL COMPARISONS OF THE THEORIES

Having assessed the logical potentials of each model, the validities of SDT and IIT are evaluated on the basis of the available empirical evidence, according to the following criteria: (1) their descriptive accuracy with respect to the behavior patterns of individuals, (2) empirical falsifications of fundamental tenets of the theories, (3) convergence of evidence across diverse methods of measurement, (4) generalizability across theoretically related tasks, and (5) process validity.

Descriptive Accuracy

Sequential decision theory. Studies using SDT do not require a perfect match between one of the ideal patterns in order to classify a child's performance. Usually, three-fourths of the responses to each problem
type must concur with the ideal pattern for that rule. Siegler (1976) noted that less than one one-millionth of the potential prediction patterns would be classifiable according to these criteria.

In one study using the balance scale, Siegler (1981) found that 80% of the kindergarteners, 88% of the third graders, 90% of the seventh graders, and 100% of the adults fit one of the sequential decision rules. Klahr and Siegler (1978) found that 77% of the 5- and 6-year-olds, 86.7% of the 9- and 10-year-olds, 93.3% of the 13- and 14-year-olds, and 100% of the 16- and 17-year-olds fit one of the sequential rules. Overall, 89.2% of their research participants fit one of the sequential rules. Ferretti et al. (1985, Experiment 1) found that among first and second graders, 64% fit Siegler's (1976) criteria for rule 1 and 15% fit rule 2. Among third and fourth graders, 37% were rule 1, 41% were rule 2, and 2% were rule 3. Among fifth and sixth graders, only 18% were rule 1, 54% were classified as rule 2, and 18% were rule 3 on the balance scale task. Overall, 83% of the children in Ferretti et al.'s (1985) study fit one of the sequential decision rules on the balance scale.

Similar levels of descriptive accuracy have been reported for SDT on other compensation tasks. Siegler (1981) found that 95% of the subjects (ages, 5 to 20 years) fit one of the sequential rules on the projection of shadows task, and 83% fit a sequential rule on the probability task. Shacklee and Mims (1981) tested 9-, 12-, and 15-year-olds, and adults on another of Inhelder and Piaget's (1958) formal operations tasks, inferring causality from information about covariation between events. Overall, one of the SDT rules fit 86% of their subjects.

Ferretti et al. (1985, Experiment 1) found developmental trends like those Siegler described on a task called the inclined plane (cf. Inhelder & Piaget, 1958, pp. 80–92). In this task, children were asked to predict which of two pinballs would travel farther after being rolled down inclined planes that vary in (a) the angle of inclination and (b) the distance from the release point to the vertex angle. Overall, one of the sequential rules fit the performance of 67% of the first through sixth graders.

Ferretti et al. (1985) also devised a method for studying the response patterns of the children that were not classifiable according to the original criteria. The strength of statistical association between an independent variable and a dependent variable can be estimated by a statistic called $\Omega^2$ (Dodd & Schultz, 1973; Hays & Winkler, 1971). Like the familiar $R^2$ in multiple regression, $\Omega^2$ yields a standardized index of association ranging from +1.00 (perfect relationship) to 0.00 (unrelated). Whereas $R^2$ represents the degree of linear relatedness between independent and dependent variables, $\Omega^2$ represents the total degree of relatedness, both linear and curvilinear (Hays & Winkler, 1971). Using the $\Omega^2$ statistic to the degree of fit between an individual's data and an external ideal pattern
(Belmont et al., 1982), Ferretti et al. (1985) compared their unclassified subjects' data to each of the ideal patterns predicted by the sequential decision theory. Based on the criterion that 75% of the variance in the child's data must fit an ideal pattern, almost all children (92%) were classified as using one of the sequential rules on the inclined plane and 96% of the children met this criterion on the balance scale. The lower percentage classified on the inclined plane was attributed to the fact that their inclined plane apparatus made the two dimensions of the task almost equal in perceptual salience, resulting in patterns that were like rule 2 or rule 3, but were too variable to meet the classification criteria for either rule.

In general, the sequential decision rules account for the behavior of more than 90% of the subjects sampled. This degree of accuracy is relatively stable across tasks and has been replicated in several laboratories. For a given task, there is a developmental trend toward increasing conformity to the sequential rules with increasing age, although the majority of preschool-aged children are classifiable on most tasks. The $\Omega^2$ method for assessing sequential rules is viewed as a substantial methodological advance over Siegler's multiple-category criteria because it provides a statistical measure of the degree of fit between any hypothesized pattern and the data of individuals for any specifiable rule pattern.

Information integration theory. Like the sequential theory, studies using the algebraic model do not require perfect correspondence between the individual's response pattern and the ideal pattern in order to consider the responses as classifiable according to the algebraic rules. For each subject, a repeated-measures analysis of variance is conducted to assess the significance of the weight and distance factors as well as the form of the function (Anderson, 1981).

Based on the requirement that the individual's $F$ ratios for both factors must be statistically significant ($p < .10$), Wilkening and Anderson (1982) reported that 76% of their 6-, 9-, and 12-year-olds, and adults were classifiable as "integrators" on the balance scale and that most of those who were classified as sequential rule 2 or rule 3 could also be classified as using an additive integration rule. It is unclear from their report whether the percentage of classifiable subjects differed according to age. Unfortunately, Wilkening and Anderson (1982) did not report what proportion of the total variance in each subject's data was associated with these additive rules.

Studies by Anderson and Cuneo (1978) and Cuneo (1982) on other tasks would appear to provide support for the information integration theory. Based on group analyses, 97.5% of the variance in 8-year-olds' ratings of liquid volume was accounted for by the main effects for height, diameter, and the linear $\times$ linear component of their interaction. The data for 11-
year-olds conformed to the multiplicative rule even more strongly. However, data for the 5-year-olds were not tested for goodness of fit, because only one dimension (height) was significant on this task. (Related findings for preschool-aged children are discussed under Falsifications, below.)

On the rectangular area task, the linear \( \times \) linear component accounted for 86.1% of the interaction between height and width for 11-year-olds, but the interaction was not significant for either the 5- or the 8-year-olds. On a subject-by-subject basis, 7 out of 10 5-year-olds had significant main effects for both the height and the width of the rectangles (Anderson & Cuneo, 1978).

Cuneo (1982) studied the effects of array length and array density on children's judgments of numerosity. In her first experiment, significant effects for both dimensions were observed only for 6 of 40 children. In another experiment, Cuneo (1982) greatly increased the statistical power by increasing the number of trials per condition from 2 (Experiment 1) to 10 (Experiment 3). Given this degree of statistical power, Cuneo (1982) found significant main effects for both array length and array density for all subjects. The fact that this degree of statistical force was required to detect any differences across levels of the second dimension makes the finding particularly unimpressive and unconvincing to knowledgeable statisticians. For example, Cronbach (1951) noted that \( F \) is dependent on the number of observations and is therefore not as good an index of the strength of a relationship as a variance ratio statistic, such as \( \Omega^2 \).

Cuneo's (1982) reported values of \( \Omega^2 \) for the linear \( \times \) linear component of the interaction term were 0.02, -0.05, -0.02, 0.07, 0.65, 0.54, and 0.72 for the 3-, 4-, 5-, 6-, 7-, and 9-year-olds, and adults, respectively. Unfortunately, Anderson and Cuneo and their colleagues have apparently always failed to report similar "proportion of total variance accounted for" statistics for the main effects. Nonetheless, \( F \) ratios for the residual component in Cuneo's (1982) single subject analyses show that only one-fourth of her children had residual \( F \) ratios greater than 1.0 and only one subject had a significant \( F \) for the residual. Performances on number quantification can be described by algebraic rules from a rather early age. However, these results can be explained equally well by a process that relies on the type of transformation performed, rather than an algebraic integration of the solution dimensions (Siegler, 1983).

Falsifications

Integration versus centration in early development. One of the outcomes that would contradict the algebraic integration theory is the finding of a significant main effect for only one of the two dimensions. This implies that only one dimension is being taken into account and therefore that no integration is in fact occurring. In contrast, the notion that a single
dimension of the task dominates young children's responding forms the central "root" of Siegler's sequential decision theory. This distinction is crucial in evaluating the two theories for describing the early development of problem-solving skills.

**Balance scale.** Surber and Gzesh (1983) used adjustment procedures to test for algebraic integration rules among 5-, 7-, and 10-year-olds, and college students. For 5-year-olds, only one of the two dimensions was significant: distance. Surber and Gzesh suggest that distance, rather than number of weights, was the dominant dimension in their study because their apparatus made the distance dimension salient relative to the weight dimension (cf. Ferretti et al., 1985, regarding the inclined plane), and because in two of their three tasks the subject was explicitly told to adjust the distance in order to change the outcome.

Ferretti et al. (1985, Experiment 2) generated prediction problems on the balance scale that distinguish the sequential rules from algebraic rules. Their test was capable of discriminating 19 distinct additive rules (weighting factors ranging from 5.0 to 0.2) and the torque (multiplying) rule from sequential rule 1, 2, or 3. (Recall that all weighted multiplying rules, including the normative torque rule, produce the same pattern on prediction tasks.) They tested children in grades 1 through 6 and found that 77% of their sample fit one of these five types of rules. One first grader and one sixth grader (3% of the total sample) were classified as using one of the 19 hypothesized adding-type rules. No children in the sample used a multiplying rule.

In contrast, 74% of the sample was classified as sequential rule 1, 2, or 3. However, detailed analyses of the rule 3 children's data on a second set of conflict problems revealed that 15 of the 18 children deviated significantly from the chance level predicted by Siegler for rule 3 (muddle through). In fact, the ideal patterns for both the multiplying and the weighted adding rules were significantly and independently related to the response patterns of children who were classified as rule 3. This last result concurs with Cuneo's (1982) two-component model (see Eq. (3)), at least for the older children in the sample. Ferretti et al. (1985) concluded that (a) there was virtually no evidence for integration rules among first-through fourth-grade children on the balance scale, (b) children classified as rule 3 by Siegler's criteria do not make random guesses on conflict problems, and (c) children classified as SDT "rule 3" show evidence of additive and multiplicative integration in their predictions on balance scale conflict problems.

**Inclined plane task.** Ferretti et al. (1985, Experiment 2) reported parallel findings on a second task, the inclined plane. Only one of the children tested was classified as using an additive integration rule and none used a multiplicative integration rule. In contrast, 41 children (48%) were clas-
sified as sequential rule 1, 2, or 3, based on Siegler's (1976) classification criteria. Twenty-two (62%) of the classifiable children were classified as rule 3 on the inclined plane task. As with the balance scale, the muddle through aspect of Siegler's rule 3 did not receive empirical support: 22/26 children who were classified as rule 3 deviated from the chance expectation in their performance on conflict items (i.e., their answers were not equally distributed across the weight-dominated, distance-dominated, and balance predictions).

Also, like their findings for the balance scale task, Ferretti et al. (1985) found that the ideal patterns for both the additive and the multiplicative integration rules were significantly and independently related to the rule 3 children's responses on inclined plane conflict problems. These results were consistent with Ferretti et al.'s interpretation of their balance scale data: algebraic rules can be detected on prediction problems using Siegler-like classification criteria, but there is virtually no evidence to support the use of integration rules by first- through sixth-grade children. Those who are classified as rule 3 do not respond randomly on conflict items. Instead, their predictions show clear influences of additive and multiplicative integration, in varying degrees (see Cuneo's (1982) Eq. (3)).

*Area volume and fullness tasks.* Anderson and Cuneo (1978) analyzed ratings data from 5-, 8-, and 11-year-olds in order to test for algebraic rules in their judgments of rectangular and triangular areas, liquid volume, and fullness. On the volume task, there was a significant main effect only for the height dimension among 5-year-olds. The 8- and 11-year-olds had significant main effects for both height and width of the glass, and the interaction indicated a multiplicative integration rule. More than 95% of the variance in the 8- and 11-year-old groups corresponded to the normative height \( \times \) surface area rule for rating liquid quantity. The 5-year-olds' volume ratings were not tested for goodness of fit because only one of the dimensions yielded a significant effect.

Unlike the volume task, ratings of the areas of rectangles showed clear-cut main effects for both height and width among the 5-year-olds as well as the older groups. The interaction was only significant for 11-year-olds. Anderson and Cuneo (1978) interpreted this result to mean that children's estimations of rectangular area involve an additive integration as early as age 5, but does not change from additive to the normative, multiplicative form until the surprisingly late age of 11.

In another study Cuneo (1982) tested 3-, 4-, 5-, 6-, 7-, and 9-year-olds', and adults' ratings of numerosity in arrays that varied in their length and density. Group data showed significant main effects for both dimensions at all age levels, but the individual analyses showed that this was a spurious result of aggregating the data of subjects who used a height-only rule with those who used a width-only rule (Cuneo, 1982).
Volume and fullness tasks. Cuneo (1983) asked 4-, 5-, 6-, and 7-year-olds to rate the volume of liquids in glasses ("amount to drink"), and then asked the same children to rate the same glasses of liquid according to how full they were. The exact same stimuli were used in both tasks, and the order of administration was balanced with one task being administered 30 days before the other. As in the original Anderson and Cuneo (1978) report, 4- and 5-year-olds' volume ratings indicated a main effect only for the height of the liquid. Six-year-olds showed main effects for both height and width and no interaction (i.e., an additive integration rule), and 7-year-olds showed a significant interaction that indicated use of the normative rule for the volume of a cylinder (height × π × r²). When the children were asked to rate how full the glasses were, "... 4-year-olds obeyed a Liquid Height - Glass Height rule [sic] . . . Judgments of 5-, 6-, and 7-year-olds obeyed the normative Liquid height/Glass height Rule." (Cuneo, 1983, p. 3). Cuneo (1980) had previously found that 3- and 4-year-olds' judgments of liquid quantity obeyed a height-only rule, even when a wide range of glass diameters were included to make the differences in this dimension more obvious. Similarly, the 4- and 5-year-olds in Cuneo's 1983 study showed no evidence of integration on one of the two tasks. Nevertheless, Cuneo concluded that "... there was no evidence of 'stimulus-bound' perceptual judgment" (Cuneo, 1983, Abstract).

Cuneo's (1983) study provides an excellent example of how the model one selects can influence the hypotheses one tests and the interpretations that one considers. Consider a glass of water. Is it necessary to suppose that the 4-year-old child (a) assesses the height of the glass, then (b) assesses the height of the liquid, then (c) computes the difference, and then (d) makes the response? This "algebraic difference" can be directly perceived as the distance from the top of the glass down to the surface of the liquid. Because this "emptiness' distance is perfectly correlated with the arithmetic "difference" between glass height and liquid height, it fits the data for the fullness task exactly as well as Cuneo's algebraic integration interpretation. This alternative explanation requires fewer assumptions regarding the complexity of internal processes. Moreover, it provides a unified explanation for the developmental pattern of results for both the volume and the fullness tasks in Cuneo's (1983) study as well as her findings on 3- and 4-year-olds' numerosity concepts. Young children did not integrate information from the two dimensions—a single dimension of the task dominated their responding in each task.

Similarly, Anderson and Cuneo (1978) themselves noted that a unidimensional "perimeter rule" fit their data for 5- and 8-year-olds' ratings of the areas of rectangles and the triangles tasks just as well as their hypothesized additive integration rules. Contrary to the repeated assertions of
Cuneo (Anderson & Cuneo, 1978; Cuneo, 1980, 1982, 1983) it appears that centration may very well play an important role in the early development of children's quantitative judgments. Furthermore, the attribution of algebraic integration to 4- and 5-year-olds on these tasks may be an artifact of the experimenters' choice of models and methods, rather than "... some peculiarity of glassware." (Anderson & Cuneo, 1978).

The great majority of evidence from both camps indicates that children do not employ a general-purpose adding rule from the very earliest ages, as has been asserted by Anderson and his collaborators. On the contrary, the evidence indicates that on many problems, young children's quantitative judgments are dominated by a single dimension.

Operativity of Algebraic Integration Rules

Three studies have been conducted to determine whether the algebraic rules defined by Anderson's (1981) methods conform to the logico-mathematical structural requirements of operativity outlined by Piaget (Inhelder & Piaget, 1958). Being able to describe a subject's data by an algebraic formula does not necessarily imply that the subject possesses a logically cohesive, algebraically organized structure.

To see if those classified as using an algebraic rule on the balance scale adjustment task understood the logical relationships between different types of compensation operations on the balance scale. Surber and Gzesh (1983) tested 5-, 7-, 10-, and 13-year-olds on three versions of the balance scale adjustment task. In one version, subjects adjusted the distance of a single weight from the fulcrum in order to counterbalance a given number of weights at a given distance on the other side of the fulcrum. In a second version, subjects adjusted the distance at which a given number of weights must be placed in order to counterbalance a given number of weights at a given distance from the other side of the fulcrum. In a third version, subjects adjusted the number of weights that must be placed at a specified distance in order to counterbalance a given amount of weight at a given distance on the other side of the scale. Surber and Gzesh found that most college students do not evidence fully reversible thinking on the balance scale. Behavior that can be described algebraically does not necessarily imply that thinking is organized algebraically (cf. Silverman & Rose, 1982).

Logical consistency of algebraic integration rules has been examined in studies of children's concept of time-speed-distance problems. Previous research had indicated that children do not truly distinguish time from speed and distance until 7 or 8 years of age, according to Piaget (1971). More detailed tests indicate that children continue to make some errors on this task as late as 12 years of age (Siegler & Richards, 1979). Using the algebraic integration approach, Wilkening (1981) showed that 5-year-old...
children inferred the amount of time required for an object to travel a given distance according to an algebraic rule in which the distance to be traveled is subtracted from the object’s speed. Thus, he concluded that children conceive of time, speed, and distance as independent conceptual units from a rather early age. However, Acredolo and Schmid (1983) presented second-, third-, fourth-, and fifth-grade children with information about starting places, speeds, and times, and then asked children to judge whether statements about the situation were logically possible. For example, “When the dog barked, the two rabbits ran away at the same speed, but the white one ran longer and farther. Is that possible?” (Acredolo & Schmid, 1983, p. 2). Only the fifth graders judged such statements accurately.

Anderson and his colleagues (e.g., Anderson & Cuneo, 1978) have argued that the quantitative judgments of children and adults constitute a “cognitive algebra,” and indeed, their data indicate that on some tasks, even preschool-aged children’s behavior can be accurately described by simple algebraic functions of the stimulus dimensions, at least when large amounts of statistical power are applied. However, tests of the logical integrity and organization of children’s quantitative judgments do not support the claim that these concepts conform to an integrated, algebraically organized, logico-mathematical structure. It would appear that the phrase cognitive algebra (Anderson, 1981) is something of a misnomer when applied to young children’s knowledge of quantitative relationships.

Convergence across Diverse Methods of Measurement

For any observable outcome, there are always a number of potential explanations. Even the most rigorous manipulative experiments are open to alternative explanations (e.g., Cuneo’s subtraction rule versus our “emptiness” rule for young children’s judgments of fullness). The first line of converging evidence for the sequential decision rules comes from analysis of children’s explanations (Siegler, 1976, 1981; Siegler & Klahr, 1982). SDT was originally developed to account for Inhelder and Piaget’s (1958) findings. The Genevans have always relied on children’s verbal explanations as their primary data. To demonstrate that the sequential decision rules were measuring the same developmental trend that Inhelder and Piaget (1958) had described, Siegler (1976) asked children to make a prediction on the balance scale, and then justify their prediction verbally. Of the 99 classifiable subjects, 84 were classified the same by their verbal explanations and their predictions. Siegler (1981) found 70–80% agreement between verbal justifications and prediction patterns on the balance scale, the projection of shadows, and the probability tasks.

With one recent exception, studies using the information integration model have not included any other measures that systematically assess
the convergent validity of the algebraic integration theory. In an ingenious study of children's time concepts, Levin et al. (1984) recently reported a developmental trend beginning with qualitative, sequential rules 1 and 2, and then shifting to the algebraic integration rules. Children were shown water running from two faucets into opaque jars at a constant line pressure. The beginning times and ending times of the two events were varied systematically. Children were asked first to predict whether the jar on the left contained more water, or the jar on the right, or if they contained the same amount of water. Then they were asked to adjust the amounts of water in the two jars by running more water into the one they thought needed more in order to equate the amounts of water.

The authors used IIT functional measurement to analyze their data, but found that a sequential decision model (cf. Siegler, 1983) fit their data for 14 of 15 subjects, whereas the (correct) additive integration rule did not fit any 7-year-olds' data. However, among 10-year-olds, the sequential rules fit only four children's data, whereas the additive IIT rules fit 10 of the remaining 11 children's data. Among 13-year-olds, SDT rules did not fit the data of any subject, whereas the additive IIT rules fit the data of 14 of 15 subjects. The overall developmental pattern parallels that reported by Ferretti et al. (1985) on the balance scale and the inclined plane tasks.

Siegler and Atlas (1981, summarized in Siegler & Klahr, 1982; Siegler & Taraban, 1986) conducted an experiment to test the "line-of-least-effort" sequential rule 4 model (Siegler, 1983) against the purely algebraic "always-compute-torque" model. Students who got virtually perfect scores on balance scale prediction tests took significantly less time to answer nonconflict problems than conflict problems, even though the same range of weight and distance values were used in both. These results support the sequential decision theory.

In another recent report, Siegler and Taraban (1986) used multiple dependent measures in order to assess children's problem-solving strategies on the balance scale. The effects of prior experience and problem difficulty on children's strategies was tested by training rule 3 children on a pretest of specific balance scale conflict problems, and then testing them on (a) nonconflict problems, (b) the conflict problems on which they had been trained, and (c) conflict problems that they had never seen before. Specific prior experience with problems produced the fastest response times, nonconflict problems produced the next fastest response times, and the untrained conflict problems produced the slowest response times. Siegler and Taraban (1986) interpreted this pattern as supporting their strength-of-association theory. This theory holds that children first search their memories for a specific response that is highly associated with the specific problem at hand, as would be the case if one had received previous training on that item. If no response is highly enough associated
with the problem, then discrete, logical decision processing is engaged, according to the sequential model (Siegler, 1976). If no definite solution is reached by logical processing, then the algebraic, quantitative solution is engaged if it is in the child's repertoire.

In general, the available studies of convergent validity of these theories have supported the sequential decision theory. SDT may have received more support to date because the studies in this area have focused on children's behavior, rather than adults' or because those who have used the algebraic model simply have not addressed this issue. It seems intuitively likely that verbal explanations from adults would be more likely to support the algebraic theory, because adults are more familiar with algebraic terms. However, in a detailed analysis of 22 college students' "think aloud" verbal protocols on balance scale problems, Hardiman, Pollatsek, and Well (1986) found only one student using an algebraic integration rule. Hardiman et al. (1986) identify a variety of heuristics used by students during the final premastery phase of development on the balance scale (Siegler's "rule 3"). However, identifying a variety of transitional rules that some college students adopt briefly during the learning process does not justify the conclusion that "It is unlikely that any simple stage analysis can characterize changes in knowledge states in more than a superficial way." (Hardiman et al., 1986, p. 80). To reject a developmental model in the absence of data from more than one age group is premature.

Cross-task Generality of the Theories

Siegler (1981) studied the degree of synchrony in children's developmental levels on proportional compensation tasks. In one experiment 3-, 4-, 5-, 8-, 12-year-olds, and adults were tested on the balance scale, the projection of shadows, and probability tasks. Only one-third of the children had the same level of rule classification on all three tasks, and only 53% had the same rule classification on two of the three tasks. The three younger groups showed 72% agreement across the three tasks, whereas the three older groups showed only 34% agreement across the three tasks. Further experimentation confirmed that the probability task followed a different developmental course than the other two tasks. Specifically, children did not progress through the hypothesized sequence of four rules on the probability task. Instead, they abruptly changed from rule 1 to the correct division rule for computing the probability of drawing the desired color of marble. In another experiment, Siegler (1981) tested 3- to 9-year-olds on conservation of number, liquid quantity, and solid quantity. Like Cuneo (1982), Siegler (1981) found some evidence for an addition rule (number = array length + array density) among children as young as 3 or 4 years. Number conservation (i.e., rule 4 on this task) preceded conservation of liquid or solid quantity for 64% of the children tested. In
contrast, more than 70% of the children used the same rule on both the liquid and the solid quantity task.

Ferretti et al. (1985, Experiment 1) assessed the similarity of elementary-school-aged children's rules on the balance scale and the inclined plane. Of those tested, 55% had the same classifications on both tasks. An additional 31% were one level more advanced on the inclined plane than on the balance scale. In a second experiment designed to detect algebraic and sequential rules, 58% of the children had the same rule classification on both tasks, replicating the degree of cross-task similarity found in their first experiment.

The data from information integration studies indicate that the use of similar algebraic rules across tasks increases with age. The studies by Anderson and Cuneo (1978) and by Cuneo (1983) showed that the youngest children used both width and height in judging areas, but only height in judging volume and fullness. However, their data clearly show that by the time children reach adolescence or early adulthood, they adhere to the multiplicative rule with extremely high consistency on tasks where it is correct (area, volume, etc.).

If the sequential rules are more appropriate for classifying younger children’s rules and the information integration model is more appropriate for older children and adults, then a system that assesses both types of rules, such as the one devised by Levin et al. (1984), might reveal higher degrees of cross-task similarity. Siegler’s (1981) finding of more cross-task similarity of rules among younger children may be due to the fact that the classification method he employed was more appropriate to the younger children’s rules and was not capable of distinguishing the varieties of integration rules that older children appear to use on these problems. By the same token, the low degree of cross-task similarity for younger children in the data of Anderson and Cuneo (1978) and Cuneo (1983) may be the result of applying a model that is more appropriate for adolescents and adults than it is for preschool- and early-elementary-school-aged children.

Process Validities

Any pattern of predictions or adjustments could be produced by more than one process. For example, Anderson and Cuneo (1978) acknowledge that their multiplication rule could very possibly be the result of a recursive adding process. Similarly, a response pattern that fits rule 3 could be produced by an additive integration process. However, the processes themselves may be revealed by the use of multiple measures that tap the hypothetical process, albeit more or less indirectly.

Butterlied (1979) provides a programmatic outline for testing process theories of intelligence. He identifies seven steps necessary for validating
theories concerning the processes responsible for behavior: (1) choose an important problem domain; (2) select criterion problems that fairly represent performance in that domain; (3) show a relationship between the problem-solving behavior and age; (4) analyze performance according to its processes within age to show the validity of the process measures; (5) show that the processes underlying performance change with age; (6) teach children to process like adults and test to see that performance also becomes more like that of instructed adults; and (7) demonstrate that this instructional manipulation can be reversed by teaching adults to process like children, thereby lowering their performance to the level of children who are taught to use the same processes.

What is the current status of research concerning the processes responsible for skill development on compensation tasks? Many of the problems listed in Table 1 are also found in introductory physics texts and other aspects of the public school core curriculum. The fair selection of problems has been highlighted by the studies of Wilkening and Anderson (1982), Levin et al. (1984), and Ferretti et al. (1985). Selecting problems that distinguish a wide variety of the feasible rules is especially important when working with a wide age range. The relationship between performance and age is well documented on many compensation tasks, whether using Siegler's rule assessment method or Anderson's functional measurement method.

The fourth step in Butterfield's program involves the analysis of processes within ages. The study of Siegler and Atlas (1981, summarized in Siegler & Taraban, 1986) showed that conflict problems required more processing time than nonconflict problems for rule 4 college students. Siegler and Taraban (1986) found that rule 3 children's responses were quicker for nonconflict problems than for conflict problems. These suggest that the time it takes to answer problems may be a useful index of the developmental level of the processing that is being applied to solve them. It also provides some initial support for the process validity of SDT. There are qualitatively different types of problems, and they elicit performances that are consistent with the sequential decision model. If rule 4 behavior is the product of a general-purpose multiplicative integration function, then the response times should vary according to the difficulty of specific combinations of values to be multiplied (cf. Sieger & Shrager, 1984), not by "problem types" as defined by the sequential decision theory. Because age itself cannot be manipulated, it is necessary to hold it constant in order to distinguish those aspects of the process that are associated with performance at any age from those aspects of the process which develop over time (Butterfield, 1979).

The relation between process and age has not been examined using response times or any other process measure on these tasks. Siegler and
Klahr (1982) included only rule 4 college students. More detailed analyses of response time patterns and their relationships with age, rule, and type of problem may provide one useful source of independent evidence for a process validation of these theories.

No previous studies have taught children to use the processing-time patterns of adults. However, Ferretti, Kerkman, and Butterfield (1982) have tried a different tactic. They have provided verbal instruction, not on what the correct answer is (cf. Siegler & Taraban, 1986) but on how to find it.

Ferretti et al. (1982) selected 6- to 10-year-olds who were rule 1 on balance scale and rule 1 or less on the inclined plane pretests. The children were taught to say rule 2: “When we have the same weights, the ones farther from the middle always go down.” (Ferretti et al., 1982, p. 5). The children were then shown how to use the rule to “win” a game based on the adjustment task. After demonstrating that they could generalize rule 2 to a prediction version of the task, the children were taught to use rule 4 on the adjustment version. Because none of the children multiplied reliably according to a pretest, the rule 4 training relied on a recursive addition process for producing the multiplicative response pattern that Siegler called rule 4. Children were required to repeat the rule, then draw a plan for the proposed solution on paper, and then test their proposed solution on a real balance scale. They were taught to write down the peg number for every weight on that peg, then add these numbers, compare the “final answer” for each side, and apply the torque rule: “The side with the biggest final answer goes down, and if the answers are the same, it [E points to the scale’s crossbar] stays level” (Ferretti et al., 1982, p. 6).

After achieving the criterion on rule 4, children were tested for generalization of the rule to the prediction version of the task. Two weeks later, these 6- to 10-year-old children were retested on the prediction version of the task with new items. All showed virtually perfect maintenance of rule 4.

A second group was matched with the first on age and scores on the balance scale and the inclined plane pretests. They were trained on the inclined plane only. The results for the inclined plane training and the balance scale were very similar.

A third group received placebo training. They were matched with the other groups for age and pretest scores, and received an amount of training on a short-term recall task equal to that received by those who were first trained on the balance scale.

After completing training on their respective first tasks, the group that had been first trained on the balance scale and the group that had been first trained on the short-term recall task were then trained on the inclined plane. Those who had been first trained on the balance scale acquired rule
4 on the inclined plane in 40% fewer trials than those who had been first trained on the placebo task. The short-term recall group did not acquire rule 4 on the inclined plane any more or less rapidly than those who were only trained on the inclined plane. The results indicated (a) that young children can be taught to use adult-like, advanced processing rules on all three tasks, (b) that training to use this adult-like strategy resulted in adult-like performance, and (c) that even young children can generalize a sophisticated problem-solving strategy across structurally isomorphic tasks. This last finding is particularly important for process validity of the sequential model. It indicates that the underlying process instructed on the balance scale aided children's acquisition on a second task, if the solution of the second task required the same underlying processes as the first.

For the tasks discussed here, no one has attempted the seventh step in Butterfield's (1979) scheme: teaching adults to process like children in order to demonstrate by reversibility that the process instruction indeed accounts for changes in performance. For example, instructing rule 4 adults to answer very rapidly on a balance scale task might produce errors on conflict problems (i.e., rule 3). Under extreme time constraints, adults' errors might become dominated by the weight dimension—a momentary developmental regression to rule 2 (or perhaps even rule 1) induced entirely by manipulating the process.

On several occasions, Anderson and his colleagues (Anderson & Cuneo, 1978, Cuneo 1982; Wilkening, 1981) have suggested mechanisms that might underlie the additive and multiplicative functions. To the best of our knowledge, no empirical evidence to support these proposed processes has been published. Perhaps because the algebraic integration theory is derived from a perceptual-psychophysical perspective, the mechanisms suggested by these authors often refer to eye movements. Eye movements constitute a potentially useful measure for validating theories of the processes responsible for performance. For example, Cuneo's (1983) subtraction rule on the liquid fullness task could be distinguished from the one-dimensional "emptiness" rule by measuring visual fixation. The subtraction rule would necessitate fixating on the bottom of the glass, whereas the emptiness rule does not. It would also seem possible to distinguish the height + width rule for rectangular area from the perimeter rule (cf. Anderson & Cuneo, 1978).

In terms of process validity, the evidence provides some support for SDT. Response times and eye movements present potentially useful measures of processing on a wide variety of tasks like those discussed in this review. Process validity cannot prove any theory to be true, but it is quite powerful as a means for ruling out theories that happen to describe performance accurately, but do not reflect the way performance is achieved.
SUMMARY

The information integration model can describe a much wider range of potential behavior patterns than the sequential decision-tree model. It also has the potential for a higher degree of precision, especially regarding the behavior of adolescent "rule 3" behavior. Logically, the sequential model is more falsifiable than the integration model, but empirically, the sequential model is less often disconfirmed, especially among preadolescents. Studies show that the sequential model describes the behavior of most preadolescents with considerable accuracy, whereas the IIT model yields accurate, determinate descriptions of adolescents' and adults' performances. A hybrid model that accounts for this developmental shift from sequential, qualitative rules to algebraic, integrative rules is likely to be the most useful for studying children's development of problem-solving skills on the balance scale and similar tasks.

In a broader arena the exercise undertaken in this review serves to illustrate the power of models to identify the latent logic behind one's reasoning, measurement methods, and research design. Further, it illustrates the potential power of empirically testing the theories responsible for those models across a range of tasks. It is trivial to say simply that each of the competing theories has some merit. It is more valuable to determine the subjects and conditions under which each is most useful, and to identify the reasons for this differential validity. In the present case, IIT seems better suited for descriptive studies of individual differences in nearmastery levels of performance, whereas the sequential theory is more useful for predicting the early levels of development, describing the transitions from one level to the next, explaining the processes that underlie performance, and guiding the design of instructional programs for promoting early development of advanced problem-solving skills.

REFERENCES


Surber, C. F. (1985a). Applications of information integration to children's social cogni-
tions. In J. B. Prior & J. D. Day (Eds.), The development of social cognition (pp. 59–94). New York: Springer-Verlag.


Received: January 8, 1987; Revised: September 29, 1987