5

Transition Processes In Quantitative Development

David Klahr

IN R. STERNBERG (ED.), MECHANISMS
OF COGNITIVE DEVELOPMENT.
SAN FRANCISCO: W.H. FREEMAN
1984.

More than 50 years have passed since Piaget first began to demonstrate that children's behavior is governed by their knowledge, rather than their lack thereof. By asking how children's knowledge generates the robust regularities in their behavior, Piaget established the theoretical and empirical precedent for the current focus on the development of knowledge. Even a casual perusal of today's developmental journals yields an unambiguous impression about the primary goals of cognitive developmental research:

1. To discover what children know. That is, to describe knowledge *states*.
2. To explain how they come to know it. That is, to describe the processes governing *transition* between states.

The two topics have received unequal attention, with the major portion of the published work addressing the former at the expense of the latter.

Perhaps that is why the editor of this book asked the contributors to "present their views on how children transit from one state to another in the course of their development." This is the fundamental question faced by cognitive developmentalists, and a fully satisfactory answer is not yet available. However, in recent years, there have been increases in our understanding of cognitive development in particular knowledge domains. These advances derive from discoveries of (1) new procedures for assessing and representing young children's knowledge states, and (2) new methods of formulating theories of transition and change.

KNOWLEDGE STATES AND TRANSITION PROCESSES

I believe that one fruitful way to approach the developmental issue "in the large" is to start with a reasonably complete account of the development of a particular, but fundamental, knowledge domain.

This chapter was supported in part by a grant from the National Science Foundation, and in part by the Spencer Foundation. The ideas presented here evolved over many years of continuing collaboration with Iain Wallace. Helpful comments on an earlier draft came from Brain MacWhinney, Robert Neches, Michael Scheier, Robert Siegler, and Robert Sternberg. Requests for reprints should be addressed to David Klahr, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pa. 15213.

In this chapter, the general themes of states and transition are imbedded in a substantive context: knowledge about quantity. Although the specific knowledge states I describe all refer to what is loosely termed "quantitative development," I will argue for the generality of the proposed transition processes.

Knowledge States

The emphasis on children's knowledge states has led to advances in the related areas of *knowledge representation* and *knowledge assessment*. By the former, I mean descriptive formalisms for characterizing the structures and processes in which children's knowledge is embedded. Knowledge representation has become a central issue in developmental psychology, warranting chapters devoted exclusively to the topic in at least two recent developmental publications (Klahr and Siegler, 1978; Mandler, 1982). The diversity of the new approaches to representation is exemplified by various descriptions of children's knowledge in terms of rules (Haith, 1980; Siegler, 1976), scripts (Nelson, 1978), skill hierarchies (Fischer, 1980), semantic nets (Gentner, 1975), grammars (Stein and Glenn, 1977), and production systems (Young and O'Shea, 1981).

These representational inventions have been accompanied by corresponding improvements in methodologies for knowledge *assessment*. Examples can be found in such diverse areas as infant quantification (Gelman, 1972a; Strauss and Curtis, 1981), scientific principles (Siegler, 1976), analogical reasoning (Sternberg, 1977), memory development (Brown, 1978; Case, 1983; Chi, 1976), and problem solving (Klahr and Robinson, 1981; Spitz and Borys, 1984), to mention but a few.

In this chapter, I will exploit both of these advances. I will draw on studies using new types of assessment of quantitative knowledge in infants and young children, and I will use the newly emergent formulation of *self-modifying production systems* to describe the developmental process that yields increasingly sophisticated quantitative knowledge. In the next few paragraphs, I will give a very brief description of what a production system is, and indicate how it can be used to describe a particular knowledge *state*. Then, I will present a state description of a "mature" form of conservation of number. This state description will subsequently be elaborated by focusing on its fundamental components: the quantifiers. That discussion will be followed by a developmental account—without detailed mechanisms. Next, I will describe some of the rudiments of self-modifying production systems; and finally, I will indicate their relevance to the question of transition.

Production-System State Descriptions. Unambiguous theories of knowledge states are a prerequisite for theories of transition, because a transition theory can be no better than a theory of what it is that is undergoing transition. During the last decade, several investigators have formulated state descriptions for knowledge domains in terms of *production systems*. The domains include seriation (Baylor and Gascon, 1974; Young, 1976), class inclusion (Klahr and Wallace, 1972), conservation and quantification (Klahr and Wallace, 1976), and balance-scale prediction (Klahr and Siegler, 1978).

A production system consists of a set of rules—called productions—written in the form of condition/action pairs; the conditions are symbolic expressions for elements of knowledge that might be present at some instant. A production system operates via a recognize/act cycle. During the recognition portion of the cycle, all the condition sides of all the productions are compared with the current contents of the immediate knowledge state. This immediate knowledge can be interpreted as primary or short-term memory (Waugh and Norman, 1965), as "M-space" (Pascual-Leone, 1970), as short-term-plus-intermediate-term memory (Bower, 1975; Hunt, 1971), or, more generally, as the currently activated portion of long-term memory.

Thus, the conditions are tests on the momentary "state of awareness' of the system. A sequence of condition elements is interpreted as a test for the simultaneous existence of that particular conjunction of individual knowledge elements. If, for a given production, all the condition elements happen to be true at some instant, we say that the production is "satisfied." If only one production is satisfied, then it "fires": the actions associated with it are taken. These actions can modify the knowledge state by adding, deleting, or changing existing elements in it, or they can correspond to interactions—either perceptual or motor—with the environment. If more than one production has its conditions satisfied, then the satisfied productions are placed into the conflict set; a conflict-resolution principle is applied; and one production fires. The act portion of the cycle executes the actions that are associated with the fired production. Then the next cycle commences. Although the human information-processing system contains an enormous number of productions, only a limited subset of long term memory is active at any one moment, and hence only a handful of productions are potentially satisfied at any instant.

Figure 5-1 shows a simple production system. P1 says that if you have a circle and a plus, replace them with a triangle. P2 says replace a triangle with a circle; P3 says that if you have two circles, replace them with a square and a plus.

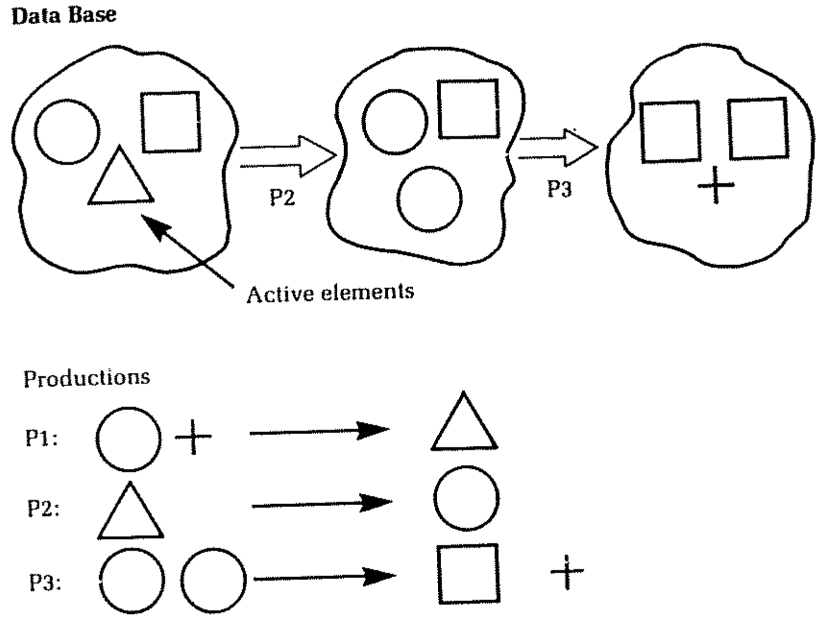

Figure 5-1
A simple production system with some active elements

If this production system were to operate on the active elements shown here—comprising a circle, a triangle, and a square—it would behave as follows. On the recognition portion of the first cycle, only P2 would have all of its conditions matched. It would fire, consuming its input and adding a circle to the knowledge state. On the next cycle, neither P1 nor P2 would be able to find a complete match, but P3 would be satisfied. It would fire, effectively replacing the two circles with a square and a plus. At this point, none of the productions would be satisfied and the system would halt.

Although one can demonstrate the logical equivalence between production systems and any of several alternative formalisms, production systems have some important practical and theoretical advantages for modeling states and transitions. The context sensitivity of production systems enables them to represent the sequential, goal-directed aspect of human cognition while maintaining the potential for interruptability. The modularity of production systems makes them particularly well suited for modeling developmental changes, because individual rules can be added, deleted, or modified without extensive consideration of the other productions.

Production systems were originally used by developmental psychologists to model children's knowledge at fixed points in time: these systems were *state descriptions*. With respect to the short time span of the performance in question (for example, a seriation task), such models were *dynamic*: they transformed one transitory knowledge state into another. However, with respect to the longer time scale of cognitive development, the earliest models were *static*: they had no capacity to change.[1] Although there might be a sequence of models representing increasingly mature knowledge states [as in the Baylor and Gascon (1974) seriation models], the early production systems had no self-modification capability.[2]

Transition

The characterization of children's behavior as "rule-governed" (cf. Haith, 1980; Siegler, 1976) is implicit in all of the knowledge mentioned earlier, and is explicitly manifested in production systems. If knowledge can be described in terms of rules, then it follows that development and learning can be conceived of in terms of rule formation, acquisition, modification, and transformation. Since such processes are themselves a form of knowledge, we can talk of *rules for learning and development*, and assume that these rules are represented in the same general form as the things that are to be learned.

Soon after the appearance of production-system models for knowledge states, psychologists began to formulate *self-modifying* production systems. The rules for modifying the production systems were stated as productions to modify other productions. These "second-generation" production systems *do* have the capacity for self-modification: they are models of transition that transform a state description from one level of performance to the next (Langley, 1980; Waterman, 1975).

Later in this chapter I will describe several important features of these models, because I believe that the self-modification mechanisms utilized in such programs provide the building blocks for an information-processing theory of developmental transitions. Prior to that, I will provide an account of the development of knowledge about quantity. The account will assume that self-modification mechanisms are available, and that they can be specified with sufficient precision to be cast as computer-simulation models. Finally, in the last section of the chapter, I will attempt to demonstrate that these mechanisms are consistent with some general characteristics of cognitive development.

Theoretical Criteria

The distinction between state theories and transition theories might imply that there is a neat temporal order to theory construction. First you describe the two states of interest; then you look at the differences between them; and finally you formulate the transition model. Of course, it is not that simple. Usually, the state descriptions that one proposes are influenced by the transition mechanism that one expects to invoke. These descriptions, in turn, affect the empirical procedures that are used to generate evidence for the existence of the processes and structures that constitute the state in question.

This seems to be the case with respect to theories of quantitative development. Claims about the temporal order or earliest manifestation of different quantitative skills are derived from a particular view of the developmental processes. Indeed, much of the empirical work described in this chapter has been chosen to support the developmental account to be presented below. One never sees transitions, only by-products of transitions; so the empirical snapshots provide a major source of support for theories of transition.

In some cases, it is difficult to choose between competing state descriptions. Even after applying such evaluative criteria as empirical fit, plausibility, parsimony, elegance, and so on, we may be left with a standoff between competing theories. It is at this point that an additional criterion, one unique to developmental psychology, can be used to evaluate theories. I call this criterion *developmental tractability*. It evaluates the extent to which the competing theories, which propose two different pairs of state descriptions for earlier and later competence, can be integrated with a transitional theory: one that can actually transform the early state into the later one. Regardless of the predictive power or elegance of a theory for a given state of knowledge, if there is no plausible mechanism that might have produced that state from some previous one, then, from the viewpoint of the developmental psychologist, such a theory is seriously deficient.

CONSERVATION OF NUMBER: A STATE DESCRIPTION

In this section, I present a state description of the basic processes involved in "mature" conservation.[3] This is the target at which a developmental account of conservation acquisition must aim. The description descends to successively more specific levels of embedding: from general rules for forming relationships among quantities (in this section), to the processes that generate quantitative information in the first place (under the next major heading). The detail

Processing steps for EC

1. $Q_i(X) \rightarrow x_i \quad Q_i(Y) \rightarrow y_i$
2. $(x_i = y_i) \longrightarrow (X \overset{Q}{=} Y)$
3. $T_p(Y) \rightarrow (Y')$
4. $(X \overset{Q}{=} Y)[T_p(Y) \rightarrow Y'] \longrightarrow (X \overset{Q}{=} Y')$

Figure 5-2
Processing steps for equivalence conservation.

is necessary because the transition processes to be proposed can be understood only in terms of the kinds of quantitative information available to them.

The classic version of the number-conservation task starts with the presentation of two distinct collections of equal amounts of discrete items (for example, two rows of beads). First the child is encouraged to establish their quantitative equality, usually by quantifying each collection independently and then noting the equivalence of the two quantifications. Then the child observes as one of the collections undergoes a transformation that changes some of its perceptual features while mantaining its numerical quantity. Finally, the child is asked to judge the relative quantity of the two collections after the transformation.

To be classified as "having conservation," the child must be able to assert the continuing quantitative equality of the two collections without resorting to a requantification and comparison after the transformation; that is, the child's response must be based not on another direct observation, but rather on the realization of the "logical necessity" for initially equal amounts to remain equal under "mere" perceptual transformations.[4] A symbolic representation of an equivalence-conservation (EC) situation is shown in Figure 5-2. Each step illustrates—in an abstract form—an essential processing phase in the standard EC task.

Construction of an Internal Quantitative Representation

The first step indicates that some *quantifier* produces an internal representation of the quantities in each of the external collections X and Y. The quantifiers are shown in their most general form as $\mathbf{Q}_i$, and their encodings as x_i. In any particular situation, one of three basic types of quantifiers may be used. The three quantifiers (to be described in more detail in the next major Section) are subitizing ($\mathbf{Q}_s$), counting ($\mathbf{Q}_c$), and estimation ($\mathbf{Q}_e$).[5] These quantifiers may produce very different internal representations of the same quantity, so information about which quantifier did the encoding is maintained in the formal notation. Thus, when collection X is quantified by $\mathbf{Q}_c$, the resultant internal symbol is indicated by x_c.

The distinction among the internal representations of quantity (or, more briefly, among *quantitive symbols*) is not merely a notational refinement. Rudimentary representations for two or three items quantified via $\mathbf{Q}_s$ may have no initial correspondence to the internal symbol generated by $\mathbf{Q}_c$ operating on the same collections. For example, an internal representation of a triangle is quite unrelated to an internal representation for the word "three." Although adults have all the required processes to detect the abstract numerical equivalence between the work representation and the triangular representation, the *initial* encodings remain, nevertheless, distinct. We can think of $\mathbf{Q}_s$ as producing something equivalent to the triangle representation, and $\mathbf{Q}_c$ as producing something equivalent to "three." It is thus clear why we want to distinguish the two possible encodings of a display of three things as x_s and x_c: they are likely to be distinct in the young child. Indeed, one of the things to be accounted for is the development of a fully generalized internal representation for number—for the "threeness of three."

Inferring External Equivalence from Internal Equality

The second processing step shown in Figure 5-2 is an inference rule that says that "if two internal quantitative symbols are equal, then their external referents are quantitatively equivalent with respect to the dimension in question." This distinction between knowledge about internal symbols and knowledge about external collections, is, like the distinction just made among quantitative symbols, a necessary part of a state description for what is ultimately to be a *developmental* theory. In adults, the inference that "three reds equal three blues implies that there are as many reds as blues" seems trivial. But it is a piece of knowledge about quantity that is not available to young children, even though they may be able to count small sets accurately. Piaget was probably the first to make the surprising discovery that young children may correctly count two equal collections, but fail to assert their quantitative equivalence:

> Fur . . . counts both sets, discovers that they have the same cardinal number, but refuses to accept that they are equivalent: "No, there are more pennies. There's one past the end." Similarly Aud counts eight pennies, says that he will be able to buy eight flowers, makes the exchange, and then cannot see that the sets are equivalent:

"There are more, because they're spread out." These cases clearly show that perception of spatial properties carries more weight than even verbal numeration (Piaget 1941/1952, p. 59)

In the somewhat more difficult situation of initial *inequivalence*, Schaeffer, Eggleston, and Scott (1974) found that young children who could correctly count four items in one set and then five in another still did not know which set had more items. In other words, children may have the ability to produce and order quantitative symbols well before they have the rule shown in step 2 of Figure 5-2. They may know that "5 comes after 4" (relational information about internal quantitative symbols), but still not know that the collection of five things is "more" than the collection of four things (external quantitative relation).

Encoding the Type of Transformation

The third step in Figure 5-2 represents an internal record that a particular type of transformation was applied to one of the collections. T_p stands for a quantity-preserving transformation.[6] The other two types of transformational classes are quantity-increasing (T_+) and quantity-decreasing (T_-). Note that the appropriate classification of a physical act depends not only on the act itself, but also on the dimension being quantified. For example, spreading is T_p with respect to number, but T_+ with respect to length and T_- with respect to density. One of the goals of a theory of conservation acquisition is to account for the formation of such transformational classes.

Applying "the" Conservation Rule

The fourth step in Figure 5-2 represents the form of knowledge that gives conservation its "logical necessity." It is a production that says, in effect, "if you know that two collections were initially equal, and that one of them underwent a quantity-preserving transformation, then you know that the transformed collection is still equal to the untransformed one." The fundamental question in the study of conservation is: where did this rule come from?

In fact, the basic question is too simple, for there is much more to be known about conservation of relationships between two quantities. There are a total of nine logically distinct conservation situations involving initial inequalities as well as equalities, and quantity-changing transformations as well as quantity-preserving ones. As shown in Figure 5-3, only seven of these can result in unambiguous outcomes. For example, in the "classic" conservation situation,

Types of transformation	Initial relations: X=Y	X>Y	X<Y
$T_p(Y) \rightarrow Y'$	X=Y' ①	X>Y' ②	X<Y' ③
$T_+(Y) \rightarrow Y'$	X<Y' ④	? ⑧	X<Y' ⑦
$T_-(Y) \rightarrow Y'$	X>Y' ⑤	X>Y' ⑥	? ⑨

Figure 5-3
The full set of possible types of conservation and nonconservation situations.

shown as Type 1, the initial relation is one of equality, the transformation is T_p, and the final relation is also equality. However, in Type 8, the initially lesser quantitiy *y* has something added to it (T_+), and the final relation between the two quantities can only be established by a final requantification.

This elaboration of types of conservation situations is important, for much of the contradictory evidence about the emergence of conservation results from a lack of attention to the particular type of conservation that is being assessed. For example, Silverman and Briga (1981) questioned Gelman's (1972b) conclusion that very young children can conserve. They evaluated the possibility that 3-year-olds solve small-number conservation problems by requantifying the final sets, rather than by using any conservation rule. Although all of their manipulations involved Type 2 and Type 3 situations, Silverman and Briga expanded their final conclusion to the broad (and ambiguous) category of "small-number conservation," asserting that their results disconfirmed Gelman's claim.

But the two studies are not comparable. In Gelman's "magic" study, the "number invariance rule" that children evidenced was that *in the absence of a transformation, number should not change.* (I will extend this argument in a later section: "Elaboration of Transformational Classes.") Although superficially Gelman's study appears to have involved an inequivalence-conservation situation (that is, a Type 2 or 3, with a surreptitious Type 8 or 9), from the children's point of view it was really a pair of identity-conservation trials: the two-item array was expected to remain a two-item array, and the three-item array a three.

It is conceivable that young children could acquire a small-number identity-conservation (IC) rule analogous to the Type 1 EC rule shown in Figure 5-3 long before they acquire the Type 2 and Type 3 rules. As for the other types, there is not much systematic empirical work, except for Siegler's (1981) discovery that the rules based on quantity-changing transformations (for example, Types 4 and 5)[7] develop prior to the analogous rules based on quantity preservation (Type 1).

QUANTIFIERS: SUBITIZING AND COUNTING

In this section I will describe some properties of both subitizing and counting in terms of their *state descriptions* at three levels: at adulthood, at about age 6, and during infancy. It is necessary to look more closely at the quantifiers before turning to a developmental account, for our theory is based on the assumption that subitizing operates in a rudimentary form very early in development.

It might be useful to summarize the structure of the presentation thus far, and to indicate where this section fits into the overall picture. Recall that in the section entitled "Theoretical Criteria," I explained that one way to assess a developmental theory was to determine the extent to which it could account for the sequence of states that were observed. With respect to quantification, then, it is important to establish just what children can do at different ages. Therefore, this section deals with some basic empirical findings and relates them to the important features of a theory of the development of quantitative skills. The section will complete the discussion of state descriptions, and will prepare us for a description of the developmental processes.

Subitizing

The phenomenology of subitizing is well known. When asked to quantify collections of fewer than four objects, adults respond very rapidly, with little conscious effort and high accuracy. The answers seem to be immediately and directly available, with no intermediate calculations. The subitizing experience is easily distinguished from the counting experience, which requires conscious attention to maintaining the correspondence between items counted and the count words, partitioning of the array into counted and uncounted items, and so on.

Subitizing in Adults. When subjects quantify random dot patterns, their reaction times (RTs)and error rates reveal a clear discontinuity at n = 3 or 4, as shown in Figure 5-4 (Chi and Klahr, 1975). For

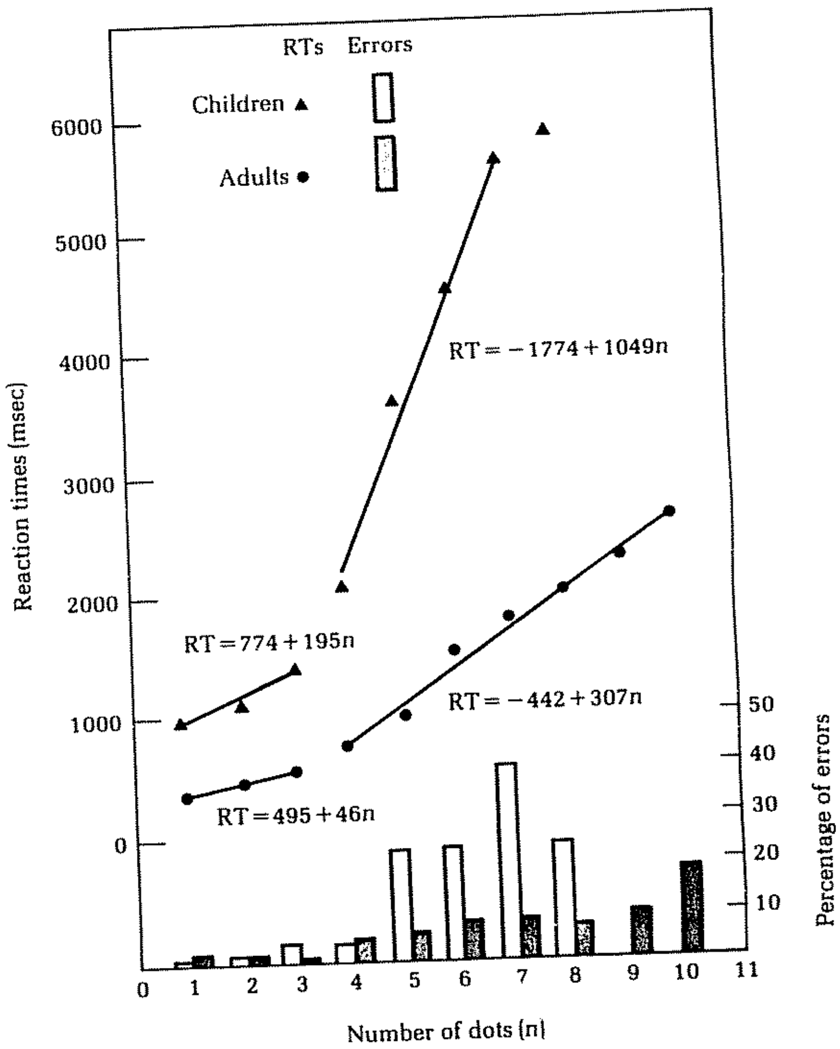

Figure 5-4
Reaction times (RTs) and error rates for children and adults in quantifying random dot patterns. From Chi and Klahr (1975, Figure 1).

adults, from n = 1 to n = 3 the *slope* of RT versus *n* is about 50 msec: RT increases about 50 msec with each additional item to be quantified. The slope changes to about 300 msec per item for *n* from 4 to 10 items. Error rates jumped from nearly zero for $n < 4$ to from 8 to 10 percent for $4 < n < 10$.[8] These results support the subjective distinction between Q_s and Q_c, but they do not support the "immediate-apprehension" view of Q_s, for it does take more time to operate on larger collections. To the extent that "pure" perceptual processes are typically characterized as highly parallel, Q_s cannot be a simple, purely perceptual process: some sequential processing must be occurring.

Furthermore, the view of Q_s as some kind of pattern recognizer cannot be reconciled with the results of studies in which subjects had to quantify complex block configurations (Chase, 1978). Using stimuli similar to the Shepard–Metzler block configurations, but varying the quantity from 1 to 10 blocks, Chase found RT patterns—for individual subjects—very similar to those in Figure 5-3. For most subjects, the upper limit of Q_s was $n = 3$, and for a few it was $n = 4$. Whereas it seems plausible that one, two, and three *dots* could be recognized as a familiar pattern, it is unlikely that people would have pattern recognizers for line drawings of small block configurations. More recently, Van Oeffelen and Vos (1982) found that people can subitize *groups* of dots as well as individual dots. They used displays of from 13 to 23 dots that could be grouped into from one to eight clusters. The displays were shown for brief periods, and the subjects had to report how many clusters of dots they saw. When the number of clusters was within the subitizing range, subjects were very accurate, but their performance deteriorated sharply when there were more than four clusters.

Thus, Q_s appears to be partially controlled by higher-order cognitive processes that determine the *target* of the quantification effort (for example, dots, cubes, dot clusters), and by some inherent limitation whereby the visual field can be rapidly segmented into only three or four perceptual chunks. Q_s is likely to be a side effect of a more general hierarchically organized perceptual process (cf. Palmer, 1977).

Subitizing in Preschoolers and Infants: Recent Evidence. Kindergarten children given the dot-quantification task produced the RT results shown by the upper curves in Figure 5-4. Also, their subjective reports for the Q_s range were similar to those of adults. We have concluded from these results that by the age of 6 years, children have both Q_c and Q_s available to them, although both processes are substantially slower than their fully mature versions.

At the time we were formulating our theory of the development of quantifiers (Klahr and Wallace, 1973), there was little evidence about the nature of infant quantification ability. The argument over whether Q_s was developmentally prior to Q_c was based in part on the developmental-tractability criterion, and in part on whether or not preschoolers used Q_c or Q_s in situations where either would

suffice (Gelman and Gallistel, 1978; Gelman and Tucker, 1975; Schaeffer, Eggleston, and Scott, 1974).

In the past few years it has been found that infants can make discriminations among small numerosities. By utilizing appropriate variations in stimulus materials, investigators have found that these discriminations are based not on brightness, total contour, extent, density, or surface area, but on number per se (Starkey and Cooper, 1980; Starkey, Spelke, and Gelman, 1980; Strauss and Curtis, 1981, 1984). Typically, infants are habituated to a particular numerosity (for example, $n = 2$). Then they are presented with post-habituation arrays of either the same numerosity or one very close to it ($n = 2$ or 3). If they dishabituate to the novel numerosity but not to the familiar one, they have demonstrated the ability to discriminate between the two quantities. Collectively, these studies demonstrate that infants as young as 4 months can discriminate 2–3 and 3–4, but not 4–5.

We now know that 4-month-old infants can perform rudimentary number discriminations long before they can have acquired the complex, socially transmitted mechanisms required by Q_c. Can we then conclude that these infants possess a precursor of Q_s? Strauss and Curtis (1981) suggest a cautious interpretation. Although their results "demonstrate that some numerosities can be discriminated by infants even though they possess no knowledge of counting," they warn that these findings "do not necessarily imply that the infant has a cognitive awareness of number and can 'represent' numerosity" (p. 1151). In order to evaulate this position, it is necessary to consider just what might be involved in the representation of numerosity.

A quantifier encodes an external stimulus into an internal quantitative symbol. The extent to which the quantifier is truly numerical depends on the extent to which it produces an internal symbol that has numerical properties—in particular, the properties of ordinality and cardinality. A full discussion of the properties of the quantitative symbols produced by the precursors of Q_c and Q_s would take us too far afield from the main theme of this chapter, which is the development of conservation rules that *utilize* Q_c and Q_s. All we need to assume about the infant representations of quantity is that they include at least cardinal information. This seems reasonable, because the habituation to a particular numerosity, when many other perceptual dimensions are varying, could only result from repeated encoding of the cardinality of the set size.

Cardinality without ordinality is only half of the story, however, and Strauss and Curtis' caution is probably prudent, for nothing from the infant labs has yet demonstrated that early quantification produces ordinal information. Several investigators, including Bullock and Gelman (1977), Estes (1976), Siegel (1971, 1977), and Silverman and Briga (1981), have demonstrated that by age 3, children can make judgments about relative numerosity. There is some disagreement about whether the relative judgments are based on number or on some correlated dimension, such as length or density (McLaughlin, 1981), but the weight of evidence appears to favor the position that preschoolers can utilize the ordinal property of number. The interesting question now concerns the extent to which *infants* can extract such information, and investigations are under way to answer this question.[9] Although we have argued elsewhere for a rudimentary quantitative representation that contains both cardinal and ordinal information,[10] for the purposes of this chapter we need to assume only that infants have a Q_s that produces reliable encodings of small cardinalities.

Counting

The complexity of counting has been convincingly demonstrated by Gelman and Gallistel's (1978) analysis of the five "counting principles" that children must understand before Q_c can produce reliable encodings of quantity. The elaboration of the principles into a computational model that can actually *do* counting is an even more impressive statement of the underlying complexity of Q_c (Greeno, Riley, and Gelman, 1981). It seems quite implausible that much of this would be available to a 4-month-old infant. Indeed, investigations of even a single component of Q_c—the ability to recite the list of number names—have revealed a protracted developmental course, typically not completed until well into the preschool years (Siegler and Robinson, 1981). It is difficult to see how Q_c could initially provide the reliable encodings necessary for the child to detect any quantitative invariants.

Nevertheless, Starkey, Spelke, and Gelman (1980) have derived from their infant studies the possibility that "infants possess a primitive form of nonverbal counting" that "may underlie (some) subitizing phenomena," and that "could share some component processes with verbal counting" (p. 9). What might this "nonverbal counting" be in infants? It certainly could not include the technology embodied in the counting principles, but it might include processes for attention deployment and symbol generation. Because such processes are common to both counting and subitizing, the infant capacity might be viewed as a precursor to either or both of the adult quantifiers. However, the similarity to Q_s derives from the finding that the upper limit of the infant discrimination seems to be three

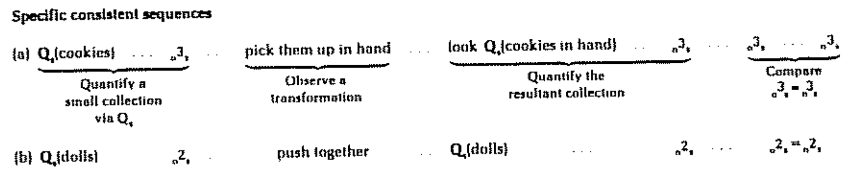

Figure 5-5
Examples of specific consistent sequences of time-line entries (a) ${}_o3_s$: old quantitative symbol for 3 objects produced by subitizing. ${}_n3_s$: new quantitative symbol for 3 objects produced by subitizing. (b) ${}_o2_s$ and ${}_n2_s$: old and new quantitative symbols for two objects produced by subitizing.

or four objects. This coincides with the break on the curves for both children's and adults' quantification times.[11]

DEVELOPMENT OF CONSERVATION OF NUMBER

Assuming, then, that initially only Q_s provides reliable encodings of small numerosities, how are the conservation rules ultimately acquired? In order to answer this question, it is necessary to describe some of the general properties of the developing information-processing system.

The Time Line

One of the fundamental capacities posited by the theory is the ability to store and analyze an encoded record of ongoing processing. This *time line* is analogous to a trace of the run of a computer program; only it is stored internally, where it is accessible to the system's basic processes for self-modification.[12] The time line and its associated analyzers can be viewed as an attempt to further specify Piaget's appealing but mysterious notion of "reflective abstraction" (Piaget, 1971); it provides a means for the system to ruminate about the efficacy of its own processing episodes. Rather than directly examining its own knowledge structures, the system regards the *results* of its actions; that is, the symbolic trace of its behavior.

Two hypothetical sequences of time-line entries relevant to quantity are shown in Figure 5-5. Entries are made in the time line at the conclusion of processing episodes. These include, in this case, simple goal-directed sensorimotor episodes, such as picking up some objects, and quantification, via Q_s, of small collections. At the early point in development represented by the sequences in Figure 5-5, the infant stores in the time line a sequence of very specific encodings of these events (as well as others, indicated by the ellipses). First, using Q_s, the infant might quantify a collection of cookies (Figure 5-5a), producing as a result a subitizing quantitative symbol for three cookies. Then he or she might encode the transformation of "picking up," followed by another execution of Q_s. Finally, the infant might notice the relation between the "old" 3 (${}_o3_s$) and the "new" 3 (${}_n3_s$) for cookies, and store that, too, in the time line.

This set of encodings—a symbolic record of quantification, transformation, requantification, and comparison of contiguous quantitative symbols—is called a *specific consistent sequence*. It is *specific* to the objects (in this example, cookies), to the number and quantifier (that is, to the particular quantitative symbol produced by Q_s), and to the transformation (picking up). It is *consistent* because both Q_s (with respect to three items) and picking up (also with respect to three items) are relatively reliable and error-free. One can imagine this particular specific sequence being entered many, many times in the time line; indeed, sufficiently often that even with uncontrolled variations in the irrelevant intervening time-line entries, the consistent pattern is detectable: the signal begins to emerge from the noise.

Other specific consistent sequences are also stored in the time line: an example of one involving different objects (dolls), a different number (2), and a different transformation (pushing together) is shown in Figure 5-5b.[13] As indicated earlier, the entries in the time line are available for inspection by the system's procedures for regularity detection. These processes look for recurrences in the stream of specific consistent sequences, and from them attempt to form *common consistent sequences*. An example of how such a common consistent sequence might be formed is diagrammed in Figure 5-6.

Generalization over Time-Line Sequences

The basic process involved is generalization. One of the three features of the specific consistent sequences from one set of experiences is compared with another, and the feature that varies is generalized while the others remain invariant. For example, Figure 5-6a shows a case in which a specific numerosity representation from Q_s and a specific transformation are discovered to be invariant with respect to the particular objects involved. The generalization produces new information in the time line: the Q_s symbol for two things, followed by spreading and requantification, followed by the discovery that

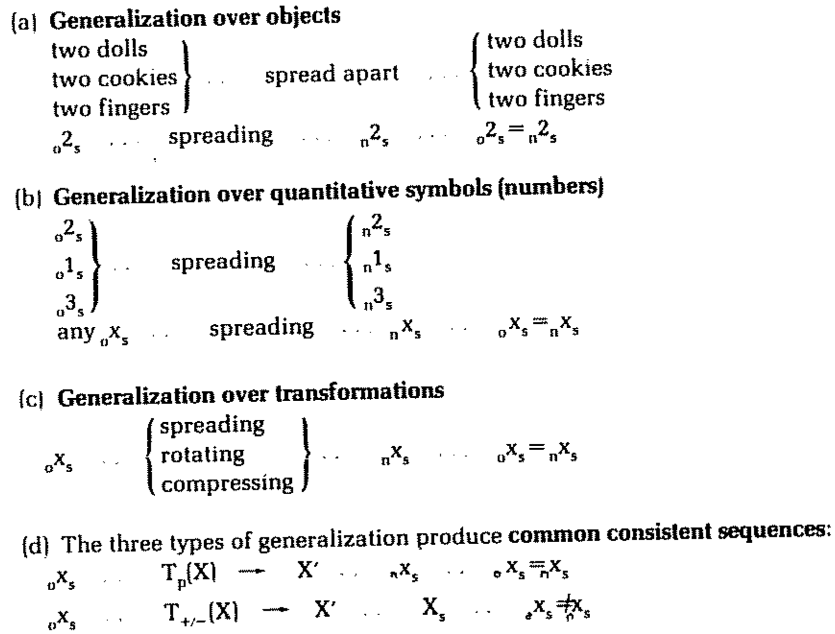

Figure 5-6
Examples of generalization over time-line entires for (a) objects, (b) quantitative symbols, and (c) transformations, to produce (d) common consistent sequences.

the second symbol is the same as the first. The time line now contains a record of activity at a slightly higher level of abstraction than before, although it is still quite limited in generality.

The next generalization shown occurs over the quantitative symbol. The minimally generalized sequences shown in Figure 5-6b are examined by the same regularity detectors that produced them. In this instance, the generalization process detects the variation in number and the constancy in transformation. In subsequent processing episodes, the system inserts information into the time line at this new level of generality: *any* Q_s symbol, followed by spreading and requantification, followed by the discovery that the first and second Q_s symbols are the same.

The partially generalized sequences shown in Figure 5-6b are then further generalized, this time over the particular transformation that is involved. If the early symbols from Q_s contain only cardinal information, then at first the only transformational generalizations will be to T_p or $T_{+/-}$ as a class. That is, the only reliable regularities

Redundancy elimination

Observe, predict, verify . . . ultimately eliminate verification

Given an IC transformation within the Q_s range, produce a "rule" or "production":

If you know then you also know

$({}_o\mathbf{x}_s)\,[T_p(X) \longrightarrow X'] \longrightarrow {}_oX_s = {}_nX_s$

Figure 5-7
Summary of rule formation via redundancy elimination for identity conservation.

will be those that include, as part of the requantification and comparison process, information that the transformation produced either the same or a different quantitative symbol.[14] With sufficiently many of these sequences, the generalization mechanisms will produce the common consistent sequences shown in Figure 5-6d: highly generalized encodings of quantification and transformation regularities (still limited to Q_s, however, since only Q_s has been producing the reliable encodings of the environment). Two forms of common consistent sequences are illustrated in Figure 5-6d. The first corresponds to quantity-preserving transformations, and the second to quantity-changing transformations.

Rule Formation via Redundancy Elimination

Common consistent sequences of the types shown in Figure 5-6d contain some redundant information. The initial Q_s is followed by the observed T_p (or $T_{+/-}$), which is always followed by another Q_s, which in turn is followed by a determination that the two quantitative symbols are the same (or are different). The developing system has a general principle of avoiding unnecessary processing—of *redundancy elimination*—and this principle is manifested in this case by processes that form a rule from the sequence. After repeated occurrences of observations that, with respect to x_s, and say, T_p, the old and the new quantities are the same, the redundancy-elimination processes form a rule about T_p. The process is sketched in Figure 5-7.

The new rule is limited to the range of *n* for which Q_s has been providing reliable encodings. There is solid evidence that the first conservations occur within the subitizing range (Cowan, 1979; Siegler, 1981).[15] Although there is nothing in the logic of the time-line processing that favors Q_s over Q_c, the added complexity of Q_c renders it an unlikely source of the requisite quantitative regularity. It has

been shown repeatedly that preschoolers can reliably count quantities well beyond their conservation range (Gelman, 1972a). If counting provided the initial basis for the formation of conservation rules, it would be hard to explain why there remains a lag of several years between reliable counting of some level of *n* and conservation of *n* things.

Suppose, instead, that the earliest conservations are specific to the internal representations produced by Q_s, as we have argued. Then the lag can be explained as a consequence of either or both of two processes. If all conservations evolve from the initial subitizing-based rules, then the lag would be due to the additional effort required for the time-line processing to detect the co-occurrence of both counting and subitizing symbols for the same quantification episode. If counting at first generates its own independent set of conservation rules, then the simple fact that counting starts later could explain the temporal discrepancy. In either case, there appears to be a period of several years when young children are using subitizing-based rules for conserving small numbers while simultaneously using counting to quantify larger collections.

Individual Variation in Conservation Acquisition

The processes that analyze the time line are postulated to be available to all infants, but the course of quantitative development is highly dependent on the particular experience of the individual. Most obvious is the expected variation in the specific items (balls, dolls, fingers, and so on) and transformations (spreading, rotating, compressing, and so on) that form the initial time-line entries. Another variation lies in the sequence of generalizations. The object, number, transformation sequence used in Figure 5-6 is not the only possible permutation of the three classes of generalizations. Depending on environmental experience, any generalization sequence might occur.

The most interesting potential variations are differences in the level of generalization at which rules may be formed. Figure 5-8 shows two orthogonal "dimensions" of change that can be effected by time-line processing: *data generalization* and *rule formation*. As illustrated in the example, the redundancy-elimination processes might produce a rule that was only partly generalized, say with respect to only 4 items and spreading, or perhaps with respect to a particular object. That is, a child might know that compressing conserves two or three pieces of candy, well before the child has the more general rule shown in Figure 5-7. These two dimensions are similar to those proposed by Newman, Riel, and Martin (1982) in their discussion of the cultural specificity of cognitive acquistions.

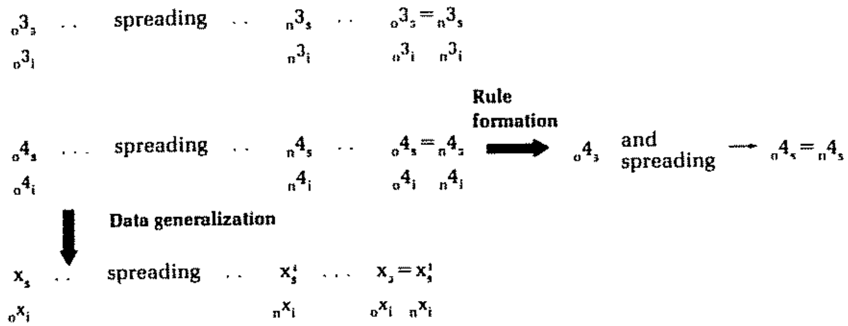

Figure 5-8
Divergent paths of data generalization and rule formation. $o3_i$: old quantitative symbol for collection of 3 items, produced by either Q_s, Q_c or Q_e.

Newman et al. note that Piaget's (1972) proposal that formal-operational reasoning is domain-specific—even in adults—is paradoxical, given the purportedly abstract nature of the psychological structures that support formal operations. Their resolution is to

> recognize that there are two orthogonal dimensions along which knowledge can be characterized. A vertical dimension, "concrete to abstract," refers to the nature of the knowledge itself, that is, the level of abstraction at which some phenomenon is understood. The horizontal dimension, "specific to general," refers to the range of contexts to which the knowledge in question applies.

It seems that Newman et al.'s "concrete to abstract" is similar to our rule-formation dimension, whereas "specific to general" corresponds to data generalization.

Expected diversity in developmental paths is not unique to the theory being proposed here. Fischer (1980), focusing on the structural rather than the procedural properties of cognitive development, makes a similar point. He argues that with respect to an individual's attainments in a particular domain (such as number concepts), "unevenness must be the rule in development" (p. 513); and that with respect to individual differences,

> for virtually every skill at every one of the levels, different individuals can take different developmental paths within a skill domain, and usually the end products of the different paths will be skills that are equivalent for most purposes. . . . [but] the different paths within a domain are often significant. (p. 514)

The process of "early" (or partially generalized) rule formation helps to explain the lack of coherence in children's "numerical understandings" (Siegler and Robinson, 1981):

> Preschoolers could have used their knowledge of counting to compare numerical magnitudes but they did not seem to. They could have used their knowledge of comparing to add numbers . . . but again they did not seem to . . . [E]arly mathematical skills may develop in relative isolation from one another. (p. 70)

Elaboration of Transformational Classes

This is not to say that no empirical predictions can be made about the appearance of conservation-related behaviors. With respect to the transformational classes, the first regularities detected would be those in which no transformation at all was encoded in the time line between two successive quantifications of the same collection. That is, the system would first come to expect that discrete quantity is constant in the absence of any transformation.[16]

It is precisely this form of "quantity permanence" that was assessed by Gelman's well-known "magic" studies (Gelman, 1972b). When 2-year-old children do not see a transformation, they expect small discontinuous quantity to remain invariant. When confronted with an array that has undergone a surreptitious numerical transformation, they show surprise.[17] Note that expecting quantitative invariance in the *absence* of an observed transformation is a simpler piece of knowledge than expecting it in the face of a specific kind of transformation (of a T_p, for example).[18] This first level in the elaboration of transformation classes is depicted in Figure 5-9a. If no transformation has been detected, then the expectation is that repeated quantification of the same collection should yield the same subitizing symbol. If any transformation *has* been observed, then there is uncertainty about what to expect. In an experiment involving pairs of sequential transformations of continuous quantity, Halford (1975) found that 4-year-old children expected almost *any* pair of transformations to produce a change in the original quantity.

The next level, shown in Figure 5-9b, involves the elaboration of transformations into either $T_{+/-}$ or non-$T_{+/-}$. Although I have been referring to the transformations that neither increase nor decrease quantity as "quantity-preserving" transformations (that is, as T_p), they are more appropriately characterized as a residual category: their meaning derives from what they *don't* do rather than from what they do. The tests in Figure 5-9b first determine whether or not any transformation has occurred. If not, the expectation is one of quantitative invariance. If there has been a transformation, a further test

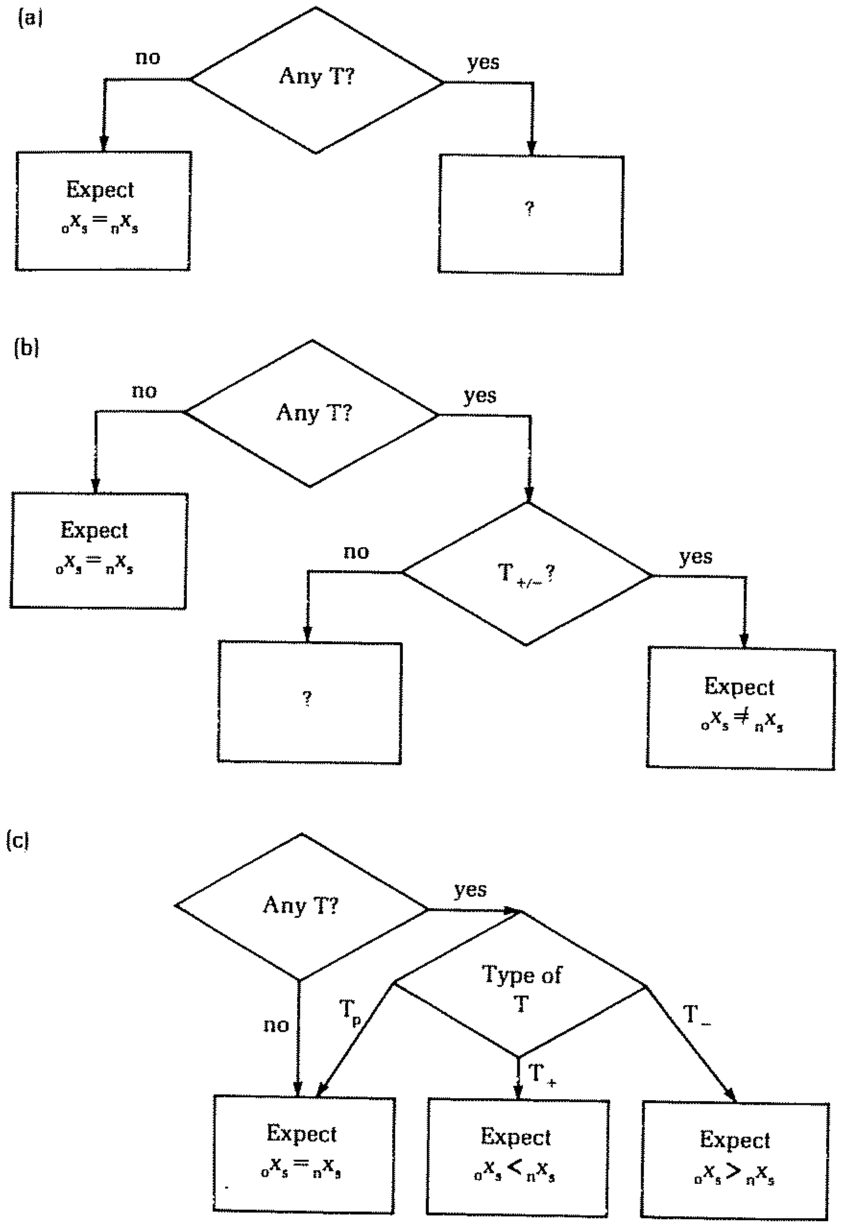

Figure 5-9
Elaboration of transformational classes.

is made to determine whether or not it is $T_{+/-}$. If it is, then a quantitative change is expected; if it is not, then no prediction is made.

Finally, in the full elaboration of transformations shown in Figure 5-9c, the specific transformation, if there is one, is directly linked with its corresponding quantitative relation. Evidence that this is a fairly late acquisition is provided by Siegler's (1981) study of the emergence of conservation rules.[19]

According to the theory presented here, "having conservation" is a consequence of acquiring the transformational knowledge illustrated in Figures 5-3 and 5-9. It does not depend on knowledge about inversion or compensation, although such knowledge may well be acquired concurrently with the acquisition of the conservation rules. This position is consistent with the results of Silverman and Rose's (1982) review of dozens of empirical studies of the relationship between compensation and conservation:

> Developmental studies support the conclusion that the ability to conserve can be attained without the ability to compensate. Further, conservation training literature points to the same conclusion. (p. 80)

MECHANISMS FOR SELF-MODIFICATION

In previous sections, I presented a state description of conservation, followed by an account of how time-line analysis leads to the new knowledge structures that support conservation. In this section, I will describe some of the mechanisms that are being used in various self-modifying production systems, and then discuss their relevance to the processes that govern the acquisition of conservation. The systems will be described at a nontechnical level, since space constraints preclude a detailed treatment of the required mechanisms.

Conflict Resolution

Before describing self-modification techniques, I should say a little more about the *conflict-resolution rules* mentioned earlier. Recall that when more than one production's conditions are satisfied, the production-system interpreter has to decide which one to "fire." Conflicts arise frequently in systems of even modest complexity, and the conflict-resolution rules are very important, especially in systems that are continually creating their own new productions. Although there are several distinct conflict-resolution rules that have been used in various production-system implementations (Anderson, Kline, and Beasley, 1980; Forgy, 1979; Langley, 1984; Newell, 1973), I will discuss only the two that are most important for self-modifying systems.

Specificity. The specificity rule selects the more specific of two otherwise equivalent productions from the conflict set. That is, if one production has conditions A and B satisfied while the other has conditions A, B, and C satisfied, then the specificity rule will choose the latter over the former.

Strength. Productions can have a *strength* associated with them. This strength is established when they are first created, and it is modified by the system in appropriate circumstances. The conflict-resolution rule that utilizes production strength will prefer the stronger of two productions in the conflict set. The stronger production is usually the one that has led to the most desirable functioning in the past. A newly created production is typically relatively weak, and a well-established one is typically strong. If the system is about to create a new production and it discovers that the same production already exists, then it will simply strengthen the existing one (Langley, 1984). Thus, initially weak productions that may be winnowed out by the strength conflict-resolution rule will, if they are repeatedly recreated, eventually become strong enough to be selected by that same process.

Discrimination

Discrimination is achieved by taking an existing production and adding more tests to its condition side. For example, if we had a production for a concept-learning task that represented the current hypothesis as

if RED and LARGE, then say YES,

and we discovered that the production was overly general, then we might want to change it to be more specific. By creating a new production having the two initial conditions plus an additional one, we might get

if RED and LARGE and TRIANGLE, then say YES.

The creation of discriminating productions usually occurs in the context of environmental feedback, and the system need only have access to the local context that caused the overly general response. Thus, it is not very difficult to decide which productions might be the most fruitful candidates for modification. Actually, the existing productions are not modified. Instead, a new production is created, and the old and new version coexist. They may follow a period during which the faulty rule still dominates the correct one. Such cases would correspond to situations in which training and experience appear to have no effect on behavior, even though the underlying correct production is getting stronger with each episode. Eventually, the conflict-resolution rules and strengthening procedures described above will ensure that the appropriate production will control behavior.

Langley (1984) proposes a system that depends entirely on discrimination as a learning and developmental mechanism. His system always starts with the most general rules possible, and then slowly builds up more and more discriminating versions of them. Langley has used this procedure to construct self-modifying production systems for concept learning, problem solving, language acquisition, and concrete operations on the balance scale.

Generalization

There are two ways to make a production more general. The first is to simply reverse the procedure for discrimination; that is, to create a new production with fewer condition elements than the one that already exists. The second way is to replace specific condition elements with variables that can be matched to any members of a class. In an algebra-learning self-modifying production system, Neves (1978) takes productions that initially refer to specific integers and creates new productions that refer just to numbers. For example, a production that starts, "if there is a 4 on the left side of the equation . . ." might get replaced with one that starts, "if there is a number . . ."

The discrimination and generalization processes can compensate for each other's excesses. Anderson and Kline (1979) describe a system that uses discrimination to recover from situations in which an overly general version of a production has been created. In the context of a concept-learning task, their system compares the situation existing when a rule is correctly applied with the situation that occurs when the same rule is incorrectly applied. Based on the differences between these two situations, the system then constructs a less general (more discriminated) version of the overly general rule.[20]

Composition

Composition is based on the idea that whenever a set of productions repeatedly fires in the same sequence, it may be possible to *combine* the set into a single production. The new production would be composed from the conditions and actions of the original set, and it would enable the system to achieve what had previously taken several recognize/act cycles in a single cycle. For example, if the system repeatedly executes the two productions

P1: if A and B, then C

and

P2: if C, then D and E,

then they could be combined into the single production

P3: if A and B, then D and E.

The next time that conditions A and B occur, the system will get to states D and E without going through the intermediate state C; it will thus get there in one step rather than two.

Composition was used by Lewis (1978) to account for two effects of practice observed with adults: speed-up and the Einstellung effect (Luchins, 1942). Its importance in a developmental context derives from the fact that, unlike discrimination and generalization, it does not depend directly on environmental feedback. The form of composition described above is based on the detection of repeated production sequences and is therefore *indirectly* dependent on what the system is encountering in the environment, but not in the sense of direct feedback about the correctness of a piece of processing. Furthermore, it is possible to do composition via a purely "syntactic" analysis of the form of the production set, without any procedural-trace information.[21]

Domains of Application

Production systems that utilize these self-modification processes have been constructed to acquire knowledge in several different contexts, including: learning to solve algebra problems (Neves, 1978); learning to solve a puzzle by repeatedly doing it (Anzai and Simon, 1979); developing efficient procedures for arithmetic (Neches, 1984) and for concept learning, schema abstraction, and language acquisition (Anderson, Kline, and Beasley, 1980; Langley, 1980); and discovering descriptive laws (Langely, 1977). Langley, Neches, Neves, and Anzai (1981) give an integrative review of this work, and offer a general characterization of the different self-modifying mechanisms that have been used.

Neches' work is especially relevant. He has constructed a system called HPM (for "heuristic procedure modification") that elaborates the basic idea of composition. HPM has a set of potentially useful ways of transforming strategies to make them more efficient. These heuristics include things such as building units, deleting unnecessary steps, saving partial results (so that they need not be recomputed), and reordering sequential steps. The heuristics operate by observing the procedural trace and determining whether or not their application is appropriate. The most extensive application of HPM has been in the context of learning to add. HPM starts with a strategy that is equivalent to children's first-acquired addition process: counting out the augend, counting out the addend, and then counting out the entire set. By applying several of its heuristics, HPM eventually modifies this strategy to the more efficient "min model." It starts the count at the larger of the two arguments, and then "counts on" the additional amount determined by the minimum argument. The most important feature of this model is that *both strategies are correct*, so the system does not change because of feedback from the environment that some error has occurred. Rather, its internal heuristics, always seeking ways to make the system more efficient, detect the inefficiencies and redundancies automatically and act on them.

Self-Modification Mechanisms and Conservation Acquisition

The time-line analysis described previously led to three major effects: the increasing generality of time-line entries, the formation of productions embodying the full set of conservation rules, and the elaboration of tests for transformational classes. How do the self-modifying procedures just presented generate the necessary effects for conservation acquisition?

Generalization over time-line sequences is accomplished by processes similar to those described earlier. Overly general rules are corrected by discrimination processes. The entries in the time line represent productions at increasingly higher levels of abstraction and generality, as the regularity-detection mechanisms examine not only the tokens for each production system's execution, but also the productions that constitute the node referred to by the token in the time line. These mechanisms then construct higher-order nodes corresponding to the common elements in the combined systems.[22] These processes are closer in flavor to Neches' HPM than to any of the other systems I have described.

Rule formation via redundancy elimination is based on procedures very similar to the composition processes explained above. It is a process that functions without any explicit external direction. The elaboration of transformational classes is accomplished via discrimination (and generalization) processes that do depend on some sort of feedback about the adequacy of the current classification of observed transformations.

I have briefly outlined some systems that can modify their productions through processes such as generalization, discrimination, composition of related conditions, and chaining of action sequences. These mechanisms provide some of the basic building blocks for a self-modifying production system that, starting with no substantive knowledge about quantity beyond the rudimentary Q_s, ultimately "acquires" conservation.

In order to construct a production system that could acquire sufficient knowledge to "have" conservation, it has been necessary to formulate increasingly powerful production-system architectures. In particular, implementation of mechanisms to generate and analyze time-line entries has added considerable complexity to the simple production-system interpreter described at the outset of this paper. It is clear that production-system approaches to developmental theory represent a strategic bet that such complexity is the necessary price of combining specificity with developmental tractability. To the best of my knowledge, there are no other equally specific proposals for how self-modification might come about.

SELF-MODIFYING SYSTEMS: LEARNING OR DEVELOPMENT?

Having proposed self-modifying production systems as viable formalisms for a theory of cognitive development, I will conclude by addressing the difficult distinction between learning and development. One frequently stated evaluation of such systems is that they may account for learning, but they certainly do not capture the "essence" of development (cf. Beilin, 1981; Neisser, 1976). But what could be the basis of such an evaluation? If we look at the many dichotomies that have been used to distinguish learning from development, the self-modifying systems appear to be more appropriately placed in the development category than in the learning category.

• *Spontaneous versus imposed.* Much of development appears to occur "on its own," without any external agent instructing, inducing, or urging the change. So, too, for some of the self-modifying systems. Time-line processing occurs continuously, and results in changes whenever the system detects the appropriate circumstances. it has the flavor of the experience-contingent spontaneity that purportedly distinguishes development from learning.

• *Qualitative versus quantitative.* This distinction has occupied philosophers and developmentalists for many years, and I can only suggest one modest clarification. Look at a program that has undergone self-modification, and ask whether the change is quantitative or qualitative. For example, in the Anzai and Simon (1979) work, it seems to me that the change from depth-first search to a recursive strategy could only be characterized as qualitative, and hence more of a developmental change than a learning one. Similarly, in Neches' (1981) model of heuristic procedure modification, the system transforms an inefficient strategy for addition (counting out the augend, counting out the addend, and then counting out the total set) into an efficient one (starting with the maximum of the two arguments and then "counting on" the other argument). It is difficult to characterize this as simply a change in which more of some pre-existing feature is added to the system: "qualitative change" seems the appropriate designation.

• *Structural reorganization versus local change.* Developmental theories, particularly those with a strong emphasis on stages (cf. Fischer, 1980), usually demand structural reorganization as a requirement for development, while viewing local changes as the province of learning. Clearly, some of the basic mechanisms in self-modifying production systems operate on a relatively local basis. Indeed, one of the great advantages of production systems is that they do not require vast systemic knowledge of the consequences of local changes. But when we begin to look carefully at changes in information-processing systems, the distinction between "local" and "structural" changes becomes blurred. Changing a few conditions in an existing production (a local change) may radically alter the firing sequence of it and all its previous successors, producing very different patterns of activation in working memory and in the time line. This in turn would result in different patterns of regularities being detected in the time line, and, ultimately, in a different set of generalizations and rules. Thus, from local changes come global effects, and from incremental modifications come structural reorganizations.[23]

• *Reflective abstraction versus practice with knowledge of results.* The systems described in this chapter constitute a very different class of models from earlier models of paired-associate learning (Feigenbaum, 1963) or concept learning (Gregg and Simon, 1967). Such models were clearly intended to account for learning in situations with externally supplied feedback about the correctness of the current state of the system. The self-modifying production systems do not necessarily get this sort of explicit feedback from the environment. Instead, many of the processes that seek pattern and regularity in the time line are completely self-contained, in the sense that they examine the trace of the system's own encodings in the absence of any clear indications of a "right" or "wrong" response. As noted earlier, the time-line processes can be viewed as a mechanization of Piaget's "reflective abstraction."

These dichotomies are not independent, nor do they exhaust the possible contrasts between learning and development. This listing should suffice, however, to show that at the level at which such contrasts are stated, there is little basis for the claim that information-processing models in general, or self-modifying production systems in particular, are inherently inadequate to capture the essence of cognitive development. This chapter is an argument for just the opposite point of view. Information-processing theories of the sort described here provide us with a powerful new language with which to write the answer to the opening question: "How do children transit from one state to another in the course of their development?"

NOTES

[1] The distinction between what is static and what is dynamic depends on the "grain" of the time interval being considered. An excellent discussion of the structure/process dichotomy and its implications for developmental psychology can be found in Newell (1973).

[2] This early state of the art led to a lot of erroneous assertions about the inherent limitations of information-processing models. For a positive and constructive response to those criticisms, see Kail and Bisanz (1983).

[3] The view of conservation acquisition presented here summarizes and extends my earlier work with J. G. Wallace (Klahr and Wallace, 1973, 1976).

[4] There are two classes of conservation tasks. In *equivalence conservation* (EC), there are two equal collections at the outset, and one of them is transformed. In *identity conservation* (IC), there is only collection, and the comparison is made between its initial and its final values (Elkind, 1967).

[5] Previously, we called these "quantification operators." Gelman has used both "estimators" (Gelman, 1972b) and "numerosity abstractors" (Gelman and Gallistel, 1978) for the same processes. Thus, the literature contains the somewhat confusing reference to subitizing and counting as two kinds of estimators. In this chapter the term "quantifier" includes the terms subitizing, counting, and estimation.

[6] Previously (Klahr and Wallace, 1976), we called this a "perceptual" transformation.

[7] More appropriately called "nonconservation rules."

[8] Although the details of these particular results—such as the exact location of the break and the linearity of the lower curve—have been subject to some strident critism (Allport, 1975), the basic conclusion is sound: subitizing and counting are different processes.

[9] See Cooper (1984) and Strauss and Curtis (1984) for the most recent work on this question:

[10] See the discussion of relative-magnitude determination in Klahr and Wallace, 1976, pp. 74–76.

[11] Our view that primitive forms of subitizing are closely tied to the specific quantitative context will be challenged if preliminary results by Starkey and Spelke (1981) turn out to be reliable. The found that for $n = 2$ or 3, infants can do cross-modal matching between simultaneous visual and sequential auditory patterns. This would argue for a highly abstract representation of small quantities.

[12] The time-line notion is more fully described in Klahr and Wallace 1976. There are formidable problems in implementing it in general form, but it has been applied in the context of arithmetic-strategy transformations (Neches, 1981), and it has been partially implemented for the conservation-acquisition domain described in this chapter (Wallace, 1979; Wallace, Klahr, and Bluff, 1984). An important feature of self-modifying production systems is their ability to represent and examine their own "procedural traces" (Langley, Neches, Neves, and Anzai, 1981), or, in our terms, to generate and analyze a time line.

[13] In order to simplify the discussion, only the identity-conservation sequences will be described at this point. The basic mechanisms for the acquisition of equivalence-conservation rules are the same as those to be described here.

[14] As the semantics of the transformations become encoded in the time line, the $\mathbf{Q}_s$ symbols can acquire the appropriate orderings as a function of the effects of transformations upon their comparison operation. A full discussion of this issue would take us too far afield at this point. All we need to assume is that the system can discriminate quantity-preserving transformations from others, be they addition or subtraction.

[15] Cowan has also concluded that the rules for identity conservation are acquired prior to those for equivalence conservation.

[16] Acredolo (1981) offers an interesting version of this position, in what he calls an "identity theory" of conservation acquisition.

[17] No such expectations would be formed with respect to continuous quantity (for example, length or density), because the en-

codings generated by rudimentary forms of Q_e are not sufficiently reliable at this point for any regularities to be detected. This is one reason why the children in the "magic" studies were not surprised by surreptitious changes in length or density. Another reason why the children were surprised by unexpected changes in number but not in length or density is that in the training phase of these studies, the children clearly established number as the discriminating feature between "winners" and "losers." Once the basis of the discrimination was removed, the children were surprised.

[18] The expectation of quantity permanence is implicit in infant habituation studies also. The increased attention to quantitative change (but not to changes in length, density, and so on) suggests that infants have similar expectations about discrete quantity not changing in the absence of some observed transformation.

[19] Siegler has proposed a similar sequence of conservation models. His are necessarily more complex because they include tests for the size of the collection and because they are descriptions of the equivalence-conservation paradigm, rather than the identity-conservation situation discussed here.

[20] See Langley (1983) for a discussion of the history of generalization and discrimination programs.

[21] Composition poses some very difficult and still unresolved problems that get to the core of the status of production systems as psychological theories. An elegant treatment of some fundamental issues can be found in Lewis (1984).

[22] See Wallace, Klahr, and Bluff (1984) for a detailed description of thse processes.

[23] There is no paradox here: poets have long been sensitive to the profound effects of local decisions:

Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

Robert Frost

REFERENCES

Acredolo, C. Acquisition of conservation: A clarification of Piagetian terminology, some recent findings, and an alternative formulation. *Human Development*, 1981, *24*, 120–137.

Allport, D. A. The state of cognitive psychology. *Quarterly Journal of Experimental Psychology*, 1975, *27*, 141–152.

Anderson, J. R., and Kline, P. J. A learning system and its psychological implications. In *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*, Tokyo, Japan, 1979.

Anderson, J. R., Kline, P. J., and Beasley, C. M., Jr. Complex learning processes. In R. E. Snow, P.-A. Federico, and W. E. Montague (Eds.), *Aptitude, learning, and instruction: Cognitive process analyses*, Vol. II. Hillsdale, N.J.: Erlbaum, 1980.

Anzai, Y., and Simon, H. A. The theory of learning by doing. *Psychological Review*, 1979, *86*, 124–140.

Baylor, G. W., and Gascon, J. An information processing theory of aspects of the development of weight seriation in children. *Cognitive Psychology*, 1974, *6*, 1–40.

Beilin, H. Piaget and the new functionalism. Address given at the Eleventh Symposium of the Piaget Society, Philadelphia, 1981.

Bower, G. H. Cognitive psychology: An introduction. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes*, Vol. I. Hillsdale, N.J.: Erlbaum, 1975.

Brown, A. L. Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology*, Vol. 1. Hillsdale, N.J.: Erlbaum, 1978.

Bullock, M., and Gelman, R. Numerical reasoning in young children: The ordering principle. *Child Development*, 1977, *48*, 427–434.

Case, R. *Intellectual development: A systematic reinterpretation*. New York: Academic Press, 1983.

Chase, W. G. Elementary information processes. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes*, Vol. 5. Hillsdale, N.J.: Erlbaum, 1978.

Chi, M. T. H. Short-term memory limitations in children: Capacity or processing deficits? *Memory and cognition*, 1976, *4*, 559–572.

Chi, M. T. H., and Klahr, D. Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology*, 1975, *19*, 434–439.

Cooper, R. G. Early number development: discovering number space with addition and subtraction. In C. Sophian (Ed.) *The origins of cognitive skills*. Hillsdale, New Jersey, Erlbaum, 1984.

Cowan, R. A reappraisal of the relation between performances of quantitative identity and quantitative equivalence conservation tasks. *Journal of Experimental Child Psychology*, 1979, *28*, 68–80.

Elkind, D. Piaget's conservation problems. *Child Development*, 1967, *38*, 15–27.

Estes, K. W. Nonverbal discrimination of more and fewer elements by children. *Journal of Experimental Child Psychology*, 1976, *21*, 393–405.

Feigenbaum, E. A. The simulation of verbal learning behavior. In E. A. Feigenbaum and J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill, 1963.

Fischer, K. W. A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 1980, *87*, 477–531.

Forgy, C. L. *The OPS74 reference manual* (Tech. rep.). Department of Computer Science, Carnegie-Mellon University, 1979.

Gelman, R. The nature and development of early number concepts. In H. W. Reese (Ed.), *Advances in child development*, Vol. 7. New York: Academic Press, 1972a.

Gelman, R. Logical capacity of very young children: Number invariance rules. *Child Development*, 1972b, *43*, 75–90.

Gelman, R., and Gallistel, C. R. *The child's understanding of number*. Cambridge, Mass.: Harvard University Press, 1978.

Gelman, R., and Tucker, M. F. Further investigations of the young child's conception of number. *Child Development*, 1975, *46*, 167–175.

Gentner, D. Evidence for the psychological reality of semantic components: The verbs of possession. In D. A. Norman and D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: W. H. Freeman and Company, 1975.

Greeno, J. G., Riley, M. S., and Gelman, R. Conceptual competence in children's counting. *Cognitive Psychology*. In press.

Gregg, L. W., and Simon, H. A. Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology*, 1967, *4*, 246–276.

Haith, M. M. *Rules that babies look by*. Hillsdale, N.J.: Erlbaum, 1980.

Halford, G. S. Children's ability to interpret transformations of a qunatity, I: An operational system for judging combinations of transformations. *Canadian Journal of Psychology*, 1975, *29*, 124–141.

Hunt, E. What kind of computer is man? *Cognitive Psychology*, 1971, *2*, 57–98.

Kail, R., and Bisanz, J. Information processing and cognitive development. In H. W. Reese (Ed.), *Advances in child development and behavior*, Vol. 17. New York: Academic Press, 1983.

Klahr, D., and Robinson, M. Formal assessment of problem solving and planning processes in preschool children. *Cognitive Psychology*, 1981, *13*, 113–148.

Klahr, D., and Siegler, R. S. The representation of children's knowledge. In H. W. Reese and L. P. Lipsitt (Eds.), *Advances in child development*, Vol. 12. New York: Academic Press, 1978.

Klahr, D., and Wallace, J. G. Class inclusion processes. In S. Farnham-Diggory (Ed.), *Information processing in children*. New York: Academic Press, 1972.

Klahr, D., and Wallace, J. G. The role of quantification operators in the development of conservation of quantity. *Cognitive Psychology*, 1973, *4*, 301–327.

Klahr, D., and Wallace, J. G. *Cognitive development: An information processing view*. Hillsdale, N.J.: Erlbaum, 1976.

Langley, P. W. BACON: A production system that discovers natural laws. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, Cambridge, Mass.: MIT Press, 1977.

Langley, P. Finding common paths as a learning mechanism. Unpublished manuscript, Carnegie-Mellon University, 1980.

Langley, P. A general theory of discrimination learning. In D. Klahr, P. Langley, and R. Neches (Eds.), *Production system: models of learning and development*. Cambridge, Mass.: MIT Press/Bradford Books, 1984.

Langley, P., Neches, R., Neves, D. M., and Anzai, Y. A domain-independent framework for procedure learning. *Policy Analysis and Information Systems*, 1981, *4*(2), 163–197.

Lewis, C. H. *Production system models of practice effects*. Ph.D. thesis, University of Michigan, 1978.

Lewis, C. Composition of productions. In D. Klahr, P. Langley, and R. Neches (Eds.), *Production system models of learning and development*. Cambridge, Mass.: MIT Press/Bradford Books, 1984.

Luchins, A. S. Mechanization in problem solving—The effects of "Einstellung." *Psychological Monographs*, 1942, *54*(248).

McLaughlin, J. A. Development of children's ability to judge relative numerosity. *Journal of Experimental Child Psychology*, 1981, *31*, 103–114.

Mandler, J. M. Representation. In P. Mussen (Ed.), *Manual of child psychology*, Vol. 2. New York: Wiley, 1982.

Neches, R. *Learning through incremental refinement of procedures* in D. Klahr, P. Langley, and R. Neches (Ed.) *Production system models of learning and development*. Cambridge, Mass.: MIT Press/Bradford Books, 1984.

Neisser, U. General, academic and artificial intelligence. In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, N.J.: Erlbaum, 1976.

Nelson, K. How children represent knowledge of their world in and out of language: A preliminary report. In R. S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Erlbaum, 1978.

Neves, D. A computer program that learns algebraic procedures by examining examples and working problems in a textbook. In *Proceedings of the Second National Conference of the Canadian Society of Computational Studies of Intelligence*, Toronto, Canada, 1978.

Newell, A. Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Newman, D., Riel, M., and Martin, L. M. W. Cultural practices and Piagetian theory: The impact of a cross-cultural research program. In D. Kuhn and J. A. Meacham (Eds.), *On the development of developmental psychology*. Basel, Switzerland: Karger, 1982.

Palmer, S. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 1977, *9*, 441–474.

Pascual-Leone, J. A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, 1970, *32*, 301–345.

Piaget, J. *The child's conception of number*. New York: Humanities Press, 1952. (C. Gattegno and F. M. Hodgson, Trans., original French edition 1941.)

Piaget, J. Intellectual evolution from adolescence to adulthood. *Human Development*, 1972, *15*, 1–12.

Piaget, T. *Biology and knowledge*. Trans. B. Walsh. Chicago: University of Chicago Press, 1971.

Schaeffer, B., Eggleston, V. H., and Scott, J. L. Number development in young children. *Cognitive Psychology*, 1974, *6*, 357–379.

Siegel, L. S. The sequence of development of certain number concepts in pre-school children. *Developmental Psychology*, 1971, *5*, 357–361.

Siegel, L. S. The cognitive basis of the comprehension and production of relational terminology. *Journal of Experimental Child Psychology*, 1977, *24*, 40–52.

Siegler, R. S. Three aspects of cognitive development. *Cognitive Psychology*, 1976, *8*(4), 481–520.

Siegler, R. S. Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, 1981, *46*(2, Serial no. 189).

Siegler, R. S., and Robinson, M. The development of numerical understandings. In H. W. Reese and L. P. Lipsitt (Eds.), *Advances in child development and behavior*, Vol. 16. New York: Academic Press, 1981.

Silverman, I. W., and Briga, J. By what process do young children solve small number conservation problems? *Journal of Experimental Child Psychology*, 1981, *32*, 115–126.

Silverman, I. W., and Rose, A. P. Compensation and conservation. *Psychological Bulletin*, 1982, *91*(1), 80–101.

Spitz, H. H., and Borys, S. V. Depth of search: How far can the retarded search through an internally represented problem space? In R. Sperber, C. McCauley, and P. Brooks (Eds.), *Learning and cognition in the mental retarded*. Baltimore, University Park Press, 1984.

Starkey, P., and Cooper, R. G., Jr. Numerosity perception in human infants. *Science*, 1980, *210*, 1033.

Starkey, P., and Spelke, E. Infants can enumerate objects and events. Paper presented at the 22nd Annual Meeting of the Psychonomic Society. Philadelphia, November, 1981.

Starkey, P., Spelke, E., and Gelman, R. Number competence in infants: Sensitivity to numeric invariance and numeric change. Paper presented at the International Conference on Infant Studies, New Haven, Conn., 1980.

Stein, N. L., and Glenn, C. G. An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Multidisciplinary approaches to discourse comprehension*. Norwood, N.J.: Ablex, 1977.

Sternberg, R. J. *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, N.J.: Erlbaum, 1977.

Strauss, M. S., and Curtis, L. E. Infant perception of numerosity. *Child Development*, 1981, *52*, 1146–1152.

Strauss, M. S. and Curtis, L. E. Development of numerical concepts in infancy. In C. Sophian (Ed.) *The origins of cognitive skills*. Hillsdale, N.J. Erlbaum, 1984.

Van Oeffelen, M. P., and Vos, P. G. Configuration effects on the enumeration of dots: Counting by groups. *Memory and Cognition*, 1982, *10*, 396–404.

Wallace, J. G. Towards a sufficient theory of transition of cognitive development. In *Proceedings of the Fifth Bienniel Congress of the International Society for the Study of Behavioral Development*, University of Lund, Sweden, 1979.

Wallace, J. G., Klahr, D., and Bluff, K. A self-modifying production system for conservation acquisition. In D. Klahr, P. Langley, and R. Neches (Eds.), *Production system models of learning and development*. Cambridge, Mass.: MIT Press/Bradford Books, 1984.

Waterman, D. Adaptive production systems. In *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*. Cambridge, Mass., Artifical Intelligence Laboratory, M.I.T., 1975.

Waugh, N. C., and Norman, D. A. Primary memory. *Psychological Review*, 1965, *72*, 89–104.

Young, R. M. *Seriation by children: An artificial intelligence analysis of a Piagetian task*. Basel, Switzerland: Birkhauser, 1976.

Young, R. M., and O'Shea, T. Errors in children's subtraction. *Cognitive Science*, 1981, *5*, 153–177.