JOURNAL TITLE:	Contributions to human development
USER JOURNAL TITLE:	Contributions to human development.
ARTICLE TITLE:	Trends in memory development research
ARTICLE AUTHOR:	Chi, M. T. H.
VOLUME:	9
ISSUE:	
MONTH:	
YEAR:	1983
PAGES:	71-107

Trends in Memory Development Research

Volume Editor
Michelene T.H. Chi, Pittsburgh, Pa.

10 figures and 4 tables, 1983

KARGER

S. Karger · Basel · München · Paris · London · New York · Tokyo · Sydney

Contr. hum. Dev., vol. 9, pp. 71–107 (Karger, Basel 1983)

A Learning Framework for Development

Michelene T.H. Chi, Ernest T. Rees
University of Pittsburgh, Pittsburgh, Pa., USA

Recently, developmental psychologists have become more aware of the importance of knowledge in the phenomena they have observed. Although traditional Piagetians often discussed knowledge, they tended to refer more to abstract abilities and logical structures. When a child acquired conservation, a concept was acquired which, while it could certainly be called knowledge, was knowledge on a plane above, and separate from, the plane of everyday factual knowledge. Just a decade ago, in the 1971 symposium on memory development, only two of the six papers made reference to knowledge. In *Corsini's* [1971, p. 231] paper, the word knowledge was not explicitly used; however, he did say, 'In a very real sense, what is coded ... is determined by the existing cognitive store at any given point in time'. In his conclusion, though, he did not place a great emphasis on 'the development of a general information base' [p. 234], listing it only as one of five possible sources of development. *Flavell* [1971, p. 273], in his discussion, did say directly, 'It has long been clear that what we know... determines what and how we perceive, or speak, or imagine, or problem solve, or predict; it is now becoming equally clear that all that knowledge ... shape(s) what and how we learn and remember'. At the conclusion of his paper, however, he referred to knowledge in the more narrow context of knowledge about storage and retrieval operations, or metaknowledge.

The fact that knowledge acquisition has not been popular as the major source of development stems from a variety of observations that seem to indicate a limitation that cannot be overcome simply by imbuing the child with more knowledge. However, more careful recent work has shown the importance of the knowledge a child already possesses to the ability to learn what is being taught. Just as it is not wise to take a course in school without the prerequisites, it is equally unwise to try to teach children knowledge for which they are unprepared. Recent advances made by developmental researchers in pinpointing exactly what knowledge a child is bringing to a task, and by researchers in cognitive science in modeling human memory and learning, necessitate that developmentalists take another look at exactly how much of development is really due to learning rather than to some kind of change in innate physical or mental structures.

Therefore, in this chapter we will present a theoretical framework in which cognitive development can be analyzed as a process of learning. By learning we mean both the acquisition and structuring of knowledge. Although it is quite likely that physical maturation sets some sort of upper limit on the prospects for learning, it seems that major developmental phenomena can largely be explained in terms of learning, especially as it relates to the structuring and restructuring of knowledge. We propose that, while learning begins with the acquisition of declarative facts, it is knowledge structures which are the internal embodiment of competence. If one possesses the concept of conservation, it is because there is a knowledge structure which represents it. If one is to make sense of incoming facts, they must be interpreted by, and stored in, existing knowledge structures. When one attains some new level of competence, it is because a new knowledge structure has been formed, perhaps by combining old ones, perhaps by creating an analog of an old one, perhaps some other way.

In the first part of this chapter we briefly summarize three well-known characteristics of development which any theory must take into account. In the second section, we discuss several kinds of explanations that have often been postulated for developmental phenomena. The third section of this paper introduces theoretical memory structures and some explicit learning mechanisms which have been postulated to operate in two general theories of learning and memory. In the fourth section we speculate on how a learning theory embodying these structures and mechanisms might explain the phenomena introduced in the first section and we also discuss some related issues.

Observed Developmental Changes

We will describe some general findings from research in both the Piagetian tradition and also in the quantitative experimental tradition.

These are major phenomena that are well documented and generally accepted by developmentalists. All theories of development should start by attempting to explain one or more of these phenomena, and if a learning theory is to be successful, it must explain them as well.

Stages

Piaget and many others [*Flavell*, 1963] have observed a general and invariant sequence in development. At the highest level are the periods which represent major changes in the child's ability to represent and interact with the world. *Piaget* called them sensory-motor, preoperational, concrete operational and formal operational. Within the overall periods, *Piaget* detected a series of stages. There are six during the sensory-motor period; however, we will be concerned with the three which span the preoperational and concrete operational periods. It is during these two periods that *Piaget* noted the acquisition of the ability to utilize such concepts as classification, ordinal relations and conservation. Within a given concept, he studied a series of related tasks such as conservation of number, conservation of substance and conservation of liquid quantity.

He described the child's performance on each task as passing through three stages. During the first two stages the child tends to center on only one dimension, and only at the third stage is the ability to decenter and consider all relevant dimensions acquired. For example, in conservation of liquid, stage I children (age 4 or 5) typically base judgements of the liquid in two containers upon the height of the liquid in each container. Stage II children (age 5 or 6) often notice the discrepancy in the alternate dimension, but cannot consolidate the information provided in both dimensions and tend to vacillate between the two in their explanations. These two stages are said to be preoperational, while stage III (age 6 or 7), the successful attainment of the conservation, is said to be concrete operational. There is a systematic progression like this within each type of conservation. When children can successfully solve all such problems, they are noted to have acquired the principle of conservation.

The original data on Piagetian tasks are quite qualitative in nature. Children are described as either successful or unsuccessful at solving a particular task. Subsequent studies, using rigorous experimental manipulations, have changed and refined *Piaget's* original notions; however, the basic finding of invariant sequences has remained. For example, *Siegler* [1981] demonstrated a sequence of increasingly more accurate understandings of the balance scale and *Siegler and Robinson* [1981] did the same for conservation of number. *Scardamalia* [1977] has shown that children of a given age cannot do combinatorial problems with N dimensions, whereas they may be able to do them with (N-1) dimensions. Similar ideas were introduced by *Halford and Macdonald* [1977], who showed that young children (age 3) cannot reproduce a checkerboard pattern that requires a code length of 2, but can reproduce one that only has a code length of 1, while the complexity of the patterns which children can reproduce increases with age.

Decalage

A child may be in different stages on tasks that seemingly require the same underlying patterns of thought. For example, although children as young as 6 may be able to perform conservation of liquid tasks, they will fail on conservation of weight tasks until age 9. *Piaget* recognized this phenomenon and referred to it as horizontal decalage, but because he believed that each stage is characterized by basic underlying structures of thought which are general and not task specific, it still poses a major obstacle to his theory. The very existence of decalage makes it clear that there must be some changes in the child that allow the later forms of each concept to be acquired.

A phenomenon analagous to decalage also appears, in non-Piagetian contexts, among adults. *Newell and Simon* [1972] noted, for example, that two problems which are isomorphs (tic-tac-toe and number scrabble) can vary considerably in difficulty. In this case, the isomorphism depends upon converting the tic-tac-toe board into a magic square with all directions adding up to 15. Likewise, adults who have successfully learned how to solve the Tower of Hanoi problem will generally not be able to apply the 'principle' underlying it to another problem isomorph, such as the Tea Ceremony [*Hayes and Simon*, 1977]. Similar examples occur in conditional reasoning in adults. College students are much better at solving equivalent conditional reasoning problems when they are couched in a familiar real-world context [*Johnson-Laird and Wason*, 1970]. Such findings suggest that decalage is not a unique characteristic of the developing child and therefore that a learning mechanism underlies it.

Memory Deficits

During the last two decades, developmental psychologists, following in the footsteps of experimental psychologists, have discovered children's deficiencies in memory abilities, particularly those pertaining to short-term memory. For example, children of age 5 can usually recall only 4 digits in a digit-span task, and children of age 7 can recall about 5 digits, and so on [*Chi*, 1976]. Related to the quantitative limit in the amount of recall is a qualitative difference in the manner with which children go about memorizing items. It is typically found that children are not as apt as adults at adopting strategies to facilitate encoding and retrieval [*Kail and Hagen*, 1977]. Even when children do use a particular strategy, their pattern of rule usage is qualitatively different from that of adults. For example, instead of rehearsing items in a cumulative fashion, they do so one by one [*Ornstein and Naus*, 1978].

Explanations for Developmental Phenomena

In this section, four types of explanations for the general developmental findings we presented in the first section will be discussed. In doing so, we will present examples of theories which embody these explanations. Since most of these theories have multiple components, they will fit under more than one of the general categories we have chosen. It should be remembered, therefore, that our purpose is to highlight important notions underlying developmental theory, not to categorize the contributions of individual authors. Also, the categories themselves are not in any way mutually exclusive; they tend to overlap in many ways, and overlap with learning theory, as well. We make our own extrapolations (when possible) in instances where the authors have not explicitly tried to explain a particular type of finding.

Capacity Increase Due to Growth

The most straightforward explanation for development is that children have to reach a certain state of physical and mental maturity before they can perform a certain task. Such theories derive in general from motor development theories, such as that of *Gesell* [1928], who documented an infant's motor capabilities with increasing age. He observed that training an infant twin on a particular motor task such as stair climbing will not result in any better performance than that of the untrained twin after a given amount of growth time, such as 9 weeks [*Gesell and Thompson*, 1929]. Although young children's central nervous systems do develop in the years after birth and theories based on 'readiness from growth' are still popular in motor learning research [*Robertson*, 1978], the current trend is to move away from that notion as an explanation for intellectual development [*Gallagher and Thomas*, 1980]. Physical growth may very well set an upper limit on developmental possibilities, but it is, as *Piaget* himself noted, the child's interaction with the environment that is of crucial importance.

However, a number of theorists have proposed that a very specific and measurable increase does take place in the capacity of short-term or working memory. The forerunner of this type of theory is *Pascual-Leone* [1970], and subsequent endorsers are *Case* [1972], *Scardamalia* [1977] and *Halford and Wilson* [1980], although *Piaget* [1928] himself also recognized such limitations. The basic notion underlying this type of theory is that performance in more complex tasks requires more items to be held in short-term memory. However, the maximum size of short-term memory is quite small, no more than seven items [*Miller*, 1956], and it is difficult to see how capacity increase alone could account for more than very simple developmental changes. Thus, this idea is usually coupled with the notion that the increased capacity allows the utilization of more complex skills, and capacity increase then becomes inseparable from increases in procedural knowledge, i.e. learning. If the skills required for a task are represented as rules, for instance, the rules for more complex tasks may require more items of information in order to execute and, perhaps, more capacity for storing intermediate results. Evidence in support of such interpretations is provided by *Scardamalia* [1977] and *Halford and Macdonald* [1977], among others. However, it is quite possible for the rules themselves to be improved in ways which allow the same short-term capacity to be utilized more efficiently, and it is difficult, if not impossible, to separate these kinds of changes from actual changes in capacity. We will describe in more detail how these changes can take place in our discussion of models of memory structures and learning processes; however, an example may make this idea clearer.

Baylor and Gascon [1974] have identified at least three basic strategies representing the three stages of development in weight seriation, as shown in table I (column 1). Stage I children have a rule or rules which allow them to compare two blocks at a time. They cannot go beyond pairs, however, and so cannot seriate the blocks at all. Stage II children essentially seriate in subseries. That is, in addition to being able to compare two blocks, they have rules which can deal with groups of 3 or 4 blocks. Stage III children finally have rules that can deal with any number of blocks. The necessary goal is to find the heaviest remaining one. Since the analysis is in

Table I. Baylor and Gascon's [1974] production system analysis of weight seriation

Stage characteristics	Mean number of items in	
	conditions	actions
Stage 1: juxtaposition of couples	1.0	1.6
Stage 2: juxtaposition of subseries	1.7	1.8
Stage 3: find heaviest	1.6	1.5

the form of production rules which have several components in the condition and action side of each rule, we have calculated the mean number of components of each. As can be seen from table I (columns 2 and 3), no obvious differences can be found. Hence, at least in this analysis, the role of the capacity of working memory does not seem to apply at the level of the size of individual rules.

Decalage provides further evidence that the invariant sequences characteristic of stages are not well explained by capacity increase alone. Stage I children can consider only one dimension; stage II children are aware of both but still use only one or the other in their explanations. Thus, it might be that at stage III, they have gained the extra capacity needed to keep both dimensions in mind at one time. However, children can be at stage I or II on one task and stage III on another at the same time. For instance, conservation of liquid is attained at age 6 or 7, conservation of weight at 9 or 10, and conservation of volume at 11 or 12. Although it is quite possible to assume that differences in the actual memory demands of such tasks are responsible for decalage, it is difficult to see how the observed increases in short-term memory ability alone could account for these differences. Also, it is possible to manipulate performance abilities on these kinds of tasks. *Gelman* [1969] has shown that children younger than the expected age can perform conservation tasks if the appropriate cues are pointed out to them. These kinds of results [see *Gelman*, 1978, for others] contradict the strict notions of stages and indicate the importance of knowledge in Piagetian tasks.

The concept of increasing short-term capacity originally came, of course, from memory research and it is easier to see how capacity changes can account for memory improvements. Basically, short-term recall is seen as a direct output of the contents of working memory [*McLaughlin*, 1963].

There may be some capacity occupied by control processes but, clearly, the larger the size of short-term memory, the greater is the ability of recall. There are problems with even this seemingly simple interpretation, however. For example, the apparent short-term memory capacity of both children and adults can be manipulated by altering the nature of the stimulus materials. *Chi* [1978] has shown that children are able to recall a greater number of items than adults when the material to be recalled is familiar to them. Similarly, adults can exhibit inferior recall when the material is not familiar to them. Thus, knowledge can interact very strongly with short-term memory abilities as well.

Thus, although the capacity of short-term memory might increase, it is not possible to explain developmental differences without postulating additional changes. Specifically, these changes are changes in knowledge, both procedural skill knowledge and general factual knowledge. Whether or not there is a capacity change, the size of short-term memory is always severely limited, and one has to learn strategies to deal ever more effectively with this limitation.

Representational Changes

Another explanation often presented for developmental differences is changes in representation, changes in the way the external environment is represented in memory. The most dramatic changes are postulated in the *mode* of representation, while more continuous and gradual changes are postulated in the *availability of memory structures*. In any event, there are two basic ways these changes can take place: maturation or learning.

Changes in the Mode of Representation. Popular conceptions of this idea are those of *Piaget* [1971] and *Bruner* et al. [1966]. The most obvious representational changes are those from an enactive (or sensory-motor) mode, occurring predominantly in infancy, to imaginal, occurring predominantly in the preoperational stage, to symbolic (or linguistic), occurring between the ages of 6 and 8. There is abundant evidence to suggest that these representations are present at these ages [*Mandler*, 1981]. Our interest here, however, centers on how the changes in representational mode can explain developmental findings of stage-like transition in problem solving, decalage and quantitative improvements in memory performance. We will focus, as an example, on the shift from imaginal to symbolic representation. Since the nature of imaginal representation is assumed to be relatively static, a child at the imaginal stage cannot represent transformation, which is required in order to solve problems such as conservation. However, the existence of decalage indicates that, while a symbolic representation might be necessary, it is not sufficient to explain success on conservation and other tasks because some of them are not attained until long after this representation is well established.

How do changes from imaginal to linguistic mode account for quantitative differences in memory performance? There are several interpretations. First, the availability of the linguistic representation probably means that there are more opportunities for the modality of the stimulus material to be compatible with the mode of representation of the stored information, thus bypassing the need to continually transform the input to a mode that would be consistent with stored information. Second, the availability of another mode of representation also permits multiple encoding, thus enhancing memory due to the duplicity of storage [*Liben*, in press; *Paivio*, 1971]. Finally, perhaps the most important reason is that having linguistic representation enhances memory performance because it facilitates the use of various verbal strategies, such as rehearsal, labelling, and so on.

The mechanism that permits the representational changes to take place often is not stated explicitly. However, *Fischer* [1980] has postulated that it is the cumulative effects of small changes in memory structures. When structures reach a certain level of complexity there is a dramatic change in the kinds of information which they can interpret and represent. For instance, when sensory-motor structures become complex enough, they can represent the relationships between motor acts and their observed consequences in one structure and, thus, become imaginal. *Kosslyn* [1978] also postulates that a large number of local changes in memory structures due to interaction with the environment cause such a change. He suggests that association, comparisons, and other mental operations initially rely on imagery. However, after frequent associations and/or comparisons, they can be stored directly. So, for example, if a child is frequently asked whether a lion or a dog is bigger, then eventually the child can answer by simply storing the proposition, 'a lion is bigger than a dog', without doing an imaginal comparison. Thus, the child's knowledge base is changed through learning. According to both *Fischer's* [1980] and *Kosslyn's* [1978] notions, changes in mode of representation come about from the accumulation of specific, localized changes in memory due to frequent exposures to environmental demand, suggesting that such changes in representational preference are not unique to children but should be demonstrable in adults learning a new domain. Thus, we would interpret changes in representation not as an explanation for development, but rather as the outcome of more fundamental processes of learning and interaction with the environment.

Another reason for the inadequacy of changes in mode as an explanation for development is simply that the level of detail is far too coarse. The changes are major ones which occur at infrequent intervals. Such major changes are suitable for explaining the major periods of development; however, many of the developmental improvements take place within the framework of one representational mode. For instance, quantitative differences in memory performance continue to occur after linguistic representation becomes dominant. In addition, the premise of the theories of *Piaget* [1971] and *Bruner* et al. [1966], that an imaginal representation is primarily static, may be wrong. Data from *Marmor* [1975], *Childs and Polich* [1979], and *Kail* et al. [1980] show that children as young as 5 are capable of performing mental rotation tasks of the *Cooper and Shepard* [1973] variety. Finally, if there is a shift in the preferred representational mode, it must occur gradually, since linguistic representation is available for children as soon as they are able to use language. Thus, the shift probably reflects a gradual change in the reliance on one sort of representation over another [*Kosslyn*, 1978], suggesting that the shift is an outcome of some more fundamental processes, rather than a cause for different levels of competence. For other arguments concerning the difficulties in explaining development by a change in mode see *Carey*, [in press] and *Mandler* [in press].

Availability of New Structures. A number of theorists have proposed a gradual increase in the complexity and sophistication of memory structures, permitting a more sophisticated sort of representation that is needed for the more complex tasks [*Fischer*, 1980; *Halford and Wilson*, 1980; *Piaget*, 1972]. Change of structures has also been called a 'fundamental reorganization of conceptual framework' [*Keil*, 1981, p. 200] or changes in the 'representational format' [*Carey*, in press]. Format level changes imply that children cannot learn a concept or solve a problem until they can represent it, which must await the availability of the new higher-level structures.

Most of these theories were postulated to explain stages and decalage. That is, the level of the knowledge structure corresponds to the level of competence. *Fischer's* [1980] ideas provide a good example. He refers to the basic units of structure as sets, a concept borrowed from mathematics.

Each set controls a certain behavior or skill such as, in a very young child, grasping. More sophisticated behaviors can be built up by combining sets into larger sets. In order to acquire a conservation, such as conservation of substance, a skill set must be built up out of existing structures in successive stages. Each such set is specific to the kind of task and each new conservation must have its own set. Thus, to acquire conservation of weight requires that the child not only realize that substance is still conserved when the clay is deformed, but also attend to its weight. The ability to attend to the weights of objects can itself be a skill set, which the child may or may not already possess, but in order to solve the more advanced conservation task successfully, the child must use both sets together. The way to do this is to *intercoordinate* them, to combine them into one larger, higher-level set which now embodies the skill of conservation of weight.

Although most theorists of structural changes do not address the findings of systematic improvements in memory performance, many of them would probably propose that the limit within each level would be the source of the deficits. *Case* [1972], for example, has conducted many memory tasks which produce performance data much like those obtained in serial recall. In one task, for example, children are shown a series of N ascending numerals (such as 5, 8, 11), one at a time. They are supposed to memorize this sequence. Then, a target numeral (such as 7) is presented, and the child is asked to insert the target in its proper place in the sequence. Since the digits are not random a memory structure which can represent ascending numbers easily can improve performance, especially as it develops room for longer sequences. Indeed, with increasing age, children are able to perform this task successfully for longer sequences, indicating that structural development may take place. Thus, as *Pascual-Leone* [1970] postulated, the level of logical structures available to the child can set limits on the apparent magnitude of the working memory.

Many ideas of structural change are reasonable. There are two interrelated problems, however: (1) the vagueness with which the processes of change are described; and (2) the lack of an independently derived criterion of what constitutes a higher-level of skill, other than children's actual competence. Perhaps the notion of new and higher-level structures can be better understood when the mechanisms that induce the emergence of these levels are more clearly elucidated. Although *Fischer* [1980] describes several transition mechanisms, he describes both the structures and the transitions rather abstractly so that it is difficult to connect them to experimental findings. For instance, he allows sets to be joined in two different ways, one producing a higher-level skill and the other producing a skill at the same level. It is not clear how one determines whether a newly joined set is a higher level of skill or on the same level. *Flavell* [1972] and *Piaget* have also provided possible transition mechanisms. In general, because all of these transition mechanisms are not stated in explicit detail, it is not possible to determine whether they are unique to development or identical to learning mechanisms.

Accessibility

Another approach to explaining development is to assume that the underlying knowledge structures do not change. What changes is the child's ability to access the 'relevant' knowledge structure. *Rozin* [1976], for example, assumes that cognitive development is the increasing ability to access or apply a skill to a wider domain of tasks and situations. Accessibility actually was popularized early in the sixties when *Flavell* [1970] discussed the notion of 'mediation deficiency'. He postulated that children often realize that they need to use a specific skill (or strategy), but simply are not able to apply it to a specific task or domain. More recently, *Brown and Campione* [1981] have stressed the notion of limited accessibility in the sense that children, even when they are experts in a particular domain, can still only access this competence in that domain and not a novel one.

The notion of limited access makes a descriptive explanation of decalage quite straightforward: The child has a rule or principle, such as conservation, but must learn to access it for each individual domain, such as liquid quantity. The same explanation can be given for children's inability to generalize strategies learned in memory tasks, and for the exceptional, but domain-specific, memory performance of expert children [*Chi*, 1978]. However, simply labelling a phenomenon does not really explain it. The concept of access simply raises further difficult questions. Is accessing knowledge structures a cognitive ability which is separate from these structures? If so, we need to know what form this ability takes and how it changes. For instance, why does the ability to access conservation take a particular sequence? An alternative explanation, which eliminates the need for a separate cognitive function, is that the observed phenomena are a consequence of the knowledge structures themselves. The following example illustrates the point.

Lawler [1981], while observing his daughter's development, documented a phenomenon which can easily be called lack of access, although he did not explicitly use that term. His daughter learned to do mental calculation involving money. At the same time, she also learned to do mental arithmetic involving pure numbers by breaking them into multiples of ten and counting up the remainders. She did not, however, connect the two techniques. For example, when asked to add 75 and 26, she said 'seventy, ninety, ninety-six, ninety-seven,...' [p. 4] and continued counting to one-hundred-one. When the question was posed in terms of money, however, she said 'That's three quarters, four, and a penny, a dollar one' [p. 4]. *Lawler* [1981] refers to these separate skills as microworlds. In both cases, she completed the sums by counting the leftover units. Thus, she had two distinct microworlds, with different conditions for their activation to accomplish what might seem to be the same logical task, and which apparently accessed the same counting skill to complete their actions. Only later did *Lawler* [1981] observe moments of insight when his daughter first noticed that she could combine her tens microworld with her money microworld.

Here we see skills that might seem to an adult to be part of the same skill, but which are actually separate. Access to the money microworld is limited to situations where money is explicitly mentioned. However, there is no need to postulate that there is some general access mechanism which causes this phenomenon. Rather it is the structure of the microworlds themselves and wider access is gained by a structural change, combining the two microworlds in some way. Thus, although lack of access is certainly a real phenomenon, it can be seen to be only a description of the effects of changing knowledge structures. Later, we will discuss several mechanisms by which knowledge structures might change through learning. One of them, *generalization*, has particular relevance to accessibility because it widens the range of application of rules. Here again, however, wider access is the result of a change in knowledge structure.

Knowledge Differences

A factor which must surely be considered in development is the simple accumulation of knowledge; older children clearly know more than younger ones [*Chi*, 1976] and, just as clearly, they obtain this knowledge through learning. Some have tried to make a distinction between theories of knowledge acquisition and theories involving structural change, labeling the former quantitative and the latter qualitative. Thus, whether the representation is a network organized into schemas [*Chi*, in press, a; *Chi and Koeske*, 1983], events [*Nelson*, 1978], scripts [*Mandler*, in press; *Nelson*, et al. this volume], production rules [*Newell*, 1973] or some other form, if new knowledge is simply added to the existing structure, the result is only a quantitative change. However, this view is probably too naive. For instance, in a rule-based representation, new knowledge means new rules and, even though they are in the same representation, they can clearly produce a qualitative change in the overall system [*Young*, 1978]. Indeed, simply adding to any structure can make it more powerful and produce an apparent qualitative change. Thus, while the qualitative-quantitative distinction is useful for our purpose of highlighting developmental explanations, it should not be taken too literally. There are two subcategories of knowledge difference theories to be discussed below.

Rule Adoption and Strategy Usage. One type of knowledge in which differences are found is procedural knowledge, knowledge of how to do things, which can be represented as rules. The changes in children's performance at different ages are explained in terms of different rules that they use at different stages. Some of the most explicit and detailed descriptions of rule use were presented by *Baylor and Gascon* [1974] on weight seriation, *Klahr and Wallace* [1972] on class inclusion and *Young* [1978] on length seriation. Essentially, each of these theories is a simulation (whether implemented on the computer or not) of the task performance, using different rules (or sets of processes) for a different level of cognitive attainment. What the simulation accomplishes is to describe the rules used by children at each stage of competence and to verify that they will indeed produce the observed behavior. We have already described *Baylor and Gascon's* [1974] analysis of the rules in table I. A similar line of reasoning is the rule assessment method of *Siegler* [1976, 1981]. Using this technique, *Siegler* was able to assess the precise rules children were using by the particular pattern of correct and erroneous responses they gave in a particular task, such as the balance scale [*Siegler,* 1981] and conservation of number [*Siegler and Robinson,* 1981].

Usage of a particular strategy or set of rules is also a common and prevalent explanation for memory improvements with age. A strategy here is usually defined as a set of processes, like rehearsal, that has been shown in the adult literature to be beneficial to remembering. And development has been shown to exhibit progressive improvement in the use of such strategies. There is an abundance of evidence and review articles on the topic [*Kail and Hagen,* 1977; *Ornstein,* 1978]. The difference between the notions here and those used in the Piagetian research is that, in the former case, we are talking about the adoption and elaboration of a particular set of rules (such as rehearsal) with increasing age, whereas in the latter case, we are talking about the adoption and use of new and more sophisticated rules at each stage of development.

Methods of assessing rule usage are very important in developmental research. At the very least, this approach allows the researcher to adequately describe in great detail the types of rules or strategies a child at a particular level is using. The intention of the rule usage approach has been to identify which components of the rules change. The implicit assumption is that once the rules of different levels are described, one can compare them to see where the differences lie and thereby understand how one rule can be transformed into another. However, transition mechanisms and transformation rules have simply not been forthcoming from this line of research. For example, it is not clear how *Baylor and Gascon's* [1974] weight seriation rules can be transformed from stage to stage, what kind of learning rules are needed, and so on. On the other hand, *Siegler's* [1981] balance scale rules do build upon each other as do *Siegler and Robinson's* [1981] number conservation rules. Thus, it may be that, at times, fairly direct learning processes act to produce modified versions of existing rules, while at other times, entirely new rules are created through the mediation of other changes in knowledge structures.

General World Knowledge. The other subcategory of knowledge which changes during development is factual, declarative knowledge of the world; clearly, children's world knowledge is less elaborated than adults'. Consequently, this gap must somehow affect children's performance in a variety of tasks. This kind of reasoning has been applied to both experimental memory tasks and Piagetian results.

Many theories have recognized the importance of world knowledge in a general way [*Brown,* 1975; *Olson,* 1973]. Some are more explicit and explain the consequences of the lack of general world knowledge in terms of chunk sizes [*Chi,* 1976; *Dempster,* 1978; *Simon,* 1972] and possibly slower access [*Chi,* 1976]. However, it was not until knowledge was explicitly manipulated that the factor of general knowledge came into prominence in developmental research [*Chi,* 1978; *Lindberg,* 1980]. Because it is difficult to define what exactly constitutes general world knowledge, experimental investigations have focused on knowledge in specific domains. Depending on how much initial knowledge the child is equipped with, researchers are able to reverse developmental trends, and/or eliminate robust developmental incapacities [*Chi,* in press, b; *Gelman,* 1978]. The importance of this knowledge component is becoming more and more convincing as greater numbers of isomorphs are appearing between the performance of a child compared to the adult versus the performance of the adult novice compared to the adult expert [*Brown,* 1982; *Chi,* in press, a].

The effect of inadequate declarative knowledge on Piagetian tasks has also been suggested by *Siegler* [1976], and more recently by *Carey* [in press], who used the same logic to postulate explanations for a variety of tasks, such as class inclusion and hypothesis testing and generation. Clearly, the child must possess some factual knowledge about the components of a task before mastering it. There is undoubtedly some effect of factual knowledge involved in decalage as well. For instance, a child who does not know what volume is will undoubtedly have great trouble mastering its conservation.

Memory Structures and Learning Mechanisms

In this section we will describe some theoretical memory structures, specifically, node-link networks, production rules and schemas. These structures have been used in a number of different theories and explanations, but we will be concerned with two which have been implemented, at least partially, on computers: ACT by *Anderson* [1976] and ASN (active structural network) by *Norman and Rumelhart* [1975]. We are interested in these two because they contain explicit learning processes to acquire, structure and restructure knowledge and because these processes have, at least to some extent, been tested and shown to be successful.

Memory Structures

Networks. Networks of nodes and links (often called propositional networks) have been very popular because they capture the associative nature of memory very effectively [*Anderson,* 1976; *Anderson and Bower,* 1973; *Collins and Loftus,* 1975; *Collins and Quillian,* 1969; *Norman and Rumelhart,* 1975; *Rumelhart* et al. 1972; *Quillian,* 1966]. Each node stands for a particular concept and the links stand for the associations or relations between nodes. Learning is the insertion of new nodes into their proper places and the acquisition of new links between existing nodes. In some models, the links can have strengths which represent the strength of association between concepts. In this case, learning can also be the strengthening of links. Alternatively, increasing strength of association might be represented by establishing multiple links between nodes [*Chi and Koeske,* 1983]. Networks are very natural representations for factual, declarative knowledge, but they can also be used to represent procedural knowledge, as we will explain in more detail shortly.

Production Rules. Production rules [*Klahr and Wallace,* 1972; *Newell,* 1973; *Newell and Simon,* 1972] can be thought of as generalized stimulus-response pairs [*Anderson,* 1976]. Each one consists of a condition side and an action side and they are often informally represented as *if=then* pairs. *If* the condition side matches the contents of short-term memory *then* the action is taken. The condition side can contain constants, which must match specific items, or variables, which can match general classes of items. The actions are generally modifications to memory. This match-action structure makes production rules ideal for representing procedural knowledge. Items can be rehearsed in short-term memory, moved from long-term to short-term memory or moved from short-term to long-term, i.e. memorized. Also goals and subgoals can be set. Since production systems are usually used for modelling cognitive activities, the processes of getting stimuli into short-term memory and controlling overt physical actions are usually ignored. In principle, however, there is no reason why they could not be modelled as well [*Klahr and Wallace,* 1972].

Production systems have several characteristics which make them quite useful for modelling human behavior and learning. First, they explicitly take the contents of short-term memory into account. This means they can handle attentional processes quite naturally. Also, because it is the contents of this memory that are generally 'seen' by such techniques as protocol analysis [*Ericsson and Simon,* 1980], comparing models with experimental data may be facillitated.

Second, they can behave very flexibly. This occurs mainly because the order of application of the rules depends upon the contents of short-term memory; one rule does not explicitly call another rule. It is only through changing the contents of this memory and setting goals that rules affect the flow of control. This style of control means that a production system can be easily interruptable. For instance, if an important piece of data enters short-term memory from the environment while some behavior is ongoing, an entirely unrelated production can match, initiating behavior appropriate to the new situation. This rapid, direct response to incoming stimuli is often referred to as stimulus-driven or bottom-up processing. Similarly, a series of productions with similar but not identical conditions can exist so that subtle differences in situations, which result in small differences in short-term memory, can cause variations in behavior. Thus, a production system can respond very quickly to important changes in a situation and very flexibly to small differences in familiar situations.

Third, production systems, although they are flexible and interruptible due to incoming data, can maintain a focus of attention and ignore irrelevant stimuli once a particular behavior is initiated. This is because, in most systems, at least one element of each condition is a goal and groups of productions which are related have the same goal. Goals and subgoals, as well as other elements, placed in short-term memory by the actions of productions strongly constrain which conditions can match. This style of processing is referred to as concept-driven or top-down. Just as it is necessary to respond to important stimuli, it is necessary to ignore unimportant inputs and maintain a focus of attention. There is abundant evidence that human behavior results from a combination of bottom-up and top-down processing, and it is very important for any model to be able to capture both at the same time.

The fourth useful characteristic of production systems is that it is relatively easy to add new rules to a system without radically altering its behavior. One reason for this characteristic is that each rule must contain a relatively small piece of knowledge. The condition side can never exceed the capacity of short-term memory, and the action side, since it usually operates on this memory as well, is constrained to be of similar size. Another reason is that rules never call each other directly, so there is no need to change other rules to call the new one or decide which rules the new one should call. Also, in many production systems, the matching process is conceived to be a parallel one in which all condition sides are tested at once, so there is no need to place a new rule in a particular location relative to the other rules. If it matches, it will be found wherever it is. The obvious importance of the ability to add new rules easily is that learning can be modelled in this way. Each new rule is a small incremental change and it is the accumulation of a lot of new rules over time that causes significant changes in the behavior of the system. In fact, new rules can be added through the action of other rules in the system [*Waterman,* 1975] so that production systems can actually learn.

Schemas. Basically, a schema is an organized unit or structure of memory that contains some body of related knowledge. Quite some time ago *Bartlett* [1932] used the schema concept to explain recall for stories.

However, most of the work on schemas and related structures, such as scripts [*Nelson* et al., this volume] and frames, is relatively recent [*Anderson* et al., 1981; *Minsky,* 1975; *Rosch and Lloyd,* 1978; *Rumelhart and Norman,* 1976; *Rumelhart and Ortony,* 1977; *Schank and Abelson,* 1977]. There is a wide variety of possible implementations of this general idea; however, only a few characteristics of schemas are important for our purposes. Schemas have slots or variables into which incoming data can fit. If enough slots are filled in a particular schema, it becomes active. As with production rules, this is often referred to as stimulus-driven or bottom-up processing. Once a schema is active, it can cause top-down processing of incoming information. Unfilled slots guide attention to relevant data, while factual information present in the schema can fill in gaps or even override inconsistent data. Although it is not necessary to specify how an individual schema is organized in order to understand how they work, it is important to note that schemas can be hierarchically organized. Two or more related schemas can be joined together into one higher level schema, and an existing schema, as it becomes more complex through learning, might develop subschemas.

To illustrate how schemas might operate, consider a child learning about a new dinosaur from a picture card. If the picture were mixed in with a group of other kinds of animals, visual features of the animal would have to fit into slots in the dinosaur schema in order to activate it; the child would have to recognize it as a dinosaur. On the other hand, if he or she knows initially what the general subject is, the dinosaur schema is already activated. In that case, the slots in this schema guide attention to various features of the picture which previous experience and learning have shown to be important. Thus, the child may look to see if it walks on two or four feet, if it has lots of sharp teeth or not, if it has a long or short neck, if it has some kind of armor and so on. Once the schema is filled in with this information, a copy of it (perhaps only partial) can be placed in long-term memory to create a specific trace of the dinosaur and its characteristics.

If the child has learned enough to discriminate different categories of dinosaurs, the overall schema may contain a set of subschemas which represent them. Thus, an upright dinosaur with a short neck and sharp teeth could fit the ferocious meat-eater schema, allowing the child to infer that it is in fact a meat-eater. This inference will enable the newly learned dinosaur to be linked with other examples from the same subschema. When a schema is active, if a piece of information (such as whether it has sharp teeth) is not provided, then facts stored in the schema can be used. Thus, if the teeth are not visible, but enough other information is present to activate the meat-eater schema, sharp teeth will be inferred. This is generally referred to as a default value.

Because schemas organize units of knowledge, they are at a higher conceptual level than production rules or networks. This level is a very useful one for many purposes because the explicit specification of individual pieces of knowledge and their interrelationships is difficult, in many cases impossible and, in many cases, such detail is simply not necessary. If a detailed specification of the structure and contents of schemas is needed, they can be implemented using either production rules or a network or a combination of both. For instance, a set of nodes in a network that is very tightly interrelated via multiple links [*Chi and Koeske,* 1983] can be a schema. Also, a group of production rules with the same goal element can be viewed as a schema, with the variables in the conditions of the rules representing the slots. The next two sections give some more examples of how schemas have been implemented.

Processes of Learning in ACT

The ACT system [*Anderson,* 1976] is designed to provide an explicit division between procedural and declarative knowledge. For this reason, it contains both production rules to represent procedures and a node-link network to represent factual knowledge. As in many network models, only part of the network is accessible, active, at any one time. A small area which is the most active represents short-term memory and is what is accessed by the condition sides of the rules. The activation spreads via the links over time in order to capture the free-associative nature of memory, and one type of learning is the strengthening of links which allows activation to flow more readily.

Although ACT is not explicitly a schema model, it is possible to build schemas into it in various ways. For instance, *Anderson* et al. [1981] give two examples of possible kinds. The first is a declarative schema which is simply an area of the network containing an organized body of factual knowledge. In fact, it is difficult to imagine how a network of knowledge could exist without identifiable schemas of this type. In this case, general production rules can access and use the schemas. Depending upon the current contents of short-term memory, the same productions can access different schemas, and the same schema can be used in different ways by different rules, to solve a problem by working forward or backward, for instance. We should note that this organization is the one we currently

prefer in our own modelling attempts [*Chi and Koeske,* 1983]. The second type of schema makes the connection between declarative and procedural knowledge more explicit. In this case, each schema has its own *schematizing productions* associated with it. The purpose of these productions is to determine if incoming stimuli fit into the slots and, if so, to activate the schema. Thus, while schemas might be modelled using the variables of production rule conditions as slots, the slots in this model are nodes with which the productions link the appropriate incoming stimuli. Once a particular schema is active, its *procedural attachments* specify procedures which can act on the items in the slots. These attachments are themselves specified very much like the main schema and might be thought of as subschemas. They specify the actual set of production rules to be used to carry out the procedure.

Acquisition of Declarative Knowledge. The accumulation of declarative knowledge must be assumed to be a fundamental learning process available to humans of all ages. Indeed, *Anderson* [1981] makes a strong case that this must be the first stage of all learning. Hence, production rules should exist which enable the system to store new declarative knowledge in the network. *Neves and Anderson* [1981] refer to this process as *declarative encoding*. Modification of the declarative knowledge, once it exists, can be accomplished in ACT by production rules which explicitly change certain structures. In fact, as we pointed out earlier, many actions of the productions constitute modifications of the declarative knowledge structure.

Acquisition of Procedural Knowledge. In ACT, procedural knowledge is acquired, after practice, through the conversion of declarative knowledge. *Anderson* [1981] refers to this transformation as *knowledge compilation* which consists of two components: *proceduralization* and *composition.* It begins with the specific knowledge necessary to perform a skill in declarative form in memory (or perhaps still in an external medium such as a textbook). General interpretive productions, productions which contain mostly general variables in their conditions, must be used to access the declarative knowledge. For example, suppose a student is learning geometry. He or she may know declaratively that in order to prove two triangles congruent, the side-angle-side (SAS) postulate is useful. This might simply be an isolated fact represented by some nodes and links or, if the student is a little further along, it might be a part of the declarative schema for SAS.

When such a problem is presented, there is, in active memory, along with various features of the problem, the very general goal of finding a way to solve it. In order to proceed, interpretive productions must find in declarative memory knowledge which has associated with it the information that it can achieve this goal. When this step is complete, the problem solver has in active memory the goal to apply the SAS postulate. Again, general productions must search declarative memory to find information, this time about how to apply SAS. This cycle continues until the problem is solved. It is very slow and halting because information must be searched out, brought into and kept in active memory at every step of the way. At times, active memory must be rehearsed to keep important information from decaying before it can be used and, at other times, the solution path attempted may overload active memory and cause important information to be lost. In addition, because very general productions are being used, they may at times retrieve declarative knowledge which seems to be useful but which, in fact, leads down blind alleys, necessitating backtracking.

Once the student has had some practice using the SAS postulate, copies of the general interpretive productions can be created with specific knowledge of the SAS postulate embedded in them. Essentially, this *proceduralization* is done by replacing variables (in both the condition and action sides) with the items from declarative memory to which they have been matching. For instance, the student might attain a new production which says: *if* you want to prove two triangles congruent, *then* try SAS. One cycle of searching declarative memory and bringing items into short-term memory is thereby eliminated, making the process faster and more reliable.

Production rules which already exist in memory can be modified by *composition*. Productions which have been applying in the same sequence whenever a skill is performed are collapsed into fewer, more powerful ones, essentially by concatenating conditions and actions of the individual rules. The productions involved might be the general interpretive productions as well as the proceduralized productions. Composition obviously results from practice, and it manifests the properties inherent in practice. First, once productions are concatenated, fewer are now needed to accomplish the same thing. Thus, the process happens faster and there is less chance of a 'wrong' rule getting into the sequence. Also, since fewer accesses of working memory are needed, there is less chance of forgetting elements and there is more short-term memory capacity available, allowing other features of the situation to be noticed which might lead to further

improvements. In sum, the skill becomes faster and more reliable, which is the result when one practices. The overall result of these two processes might be to create a single production rule which automatically causes the proper correspondences between sides and angles to be made whenever a triangle congruence problem is encountered.

Further practice continues to improve performance through the general process of *tuning*. *Generalization* and *specialization* can cause more accurate versions of productions to be created and further composition can also occur. To generalize, the range of items which will match the conditions is increased, such as by replacing constants with variables. For example, suppose that a child has the following production:

'*If* the goal is to remember, and there is a string of digits, *then* repeat them one by one'.

Generalization can occur by replacing the condition 'a string of digits' by the variable 'a string of items'. Specialization works in reverse; general variables are replaced with more specific items. The net result of all of these processes for our geometry student might be to create a rule whose condition side is specific to just those triangle congruence problems with features suitable for SAS. Thus, the student may be able to recognize how to do such problems without any apparent conscious effort.

Processes of Learning in ASN

Unlike ACT, ASN uses only a node-link network to represent all knowledge. This network is organized into schemas with the characteristics we noted earlier, slots and a hierarchical structure. *Rumelhart and Norman* [1976] propose that these schemas allow one to organize, expand, understand and store inputs, as in our dinosaur example. Similarly they allow one to interpret memories when they are recalled. They also can control actions [*Rumelhart and Norman*, 1982].

The unitary representation allows the same knowledge to be both procedural and declarative at the same time. Thus, a dinosaur schema may consist of the node 'dinosaur' linked to the node 'ferocious meat-eater', with links from this node to 'upright', 'lots of sharp teeth' and 'short neck'. This structure, which is probably part of a more general dinosaur schema, could be accessed as declarative knowledge. On the other hand, this same structure can represent a set of instructions for determining if a dinosaur is a ferocious meat-eater. ASN contains a general interpreter which can trace through a set of links like this and cause the general actions of checking the teeth, and so on, to take place.

Rumelhart and Norman [1976] have proposed two general types of learning based upon schemas. Like *Anderson*, they believe acquisition of declarative facts comes first. They call it *accretion*. Existing schemas are used to store items in their proper places in the network. The other type of learning consists of several processes which operate at the schema level. A *tuning* process fairly similar to that in ACT can occur. *Generalization* here occurs within a schema by increasing the range of items which will fit slots, and *specialization* happens by decreasing their range or by replacing them with specific default values.

Schemas can also produce new schemas, a process *Rumelhart and Norman* [1976] call *restructuring*. This can be accomplished in two ways. *Patterned generation* is the creation of a new schema by modifying an old one. Thus, if no existing schemas fit a given situation well enough, a new one is created using the one that fits the best. This process can be viewed as learning by analogy [*Rumelhart and Norman*, 1981]. *Schema induction* is the creation of a new schema from two or more older ones which tend to occur together. This process is characterized as a sort of 'aha' phenomenon involving insight, and presumably occurs rarely.

Summary of Learning Mechanisms

There are many similarities in the learning mechanisms proposed for these two models. An accumulation of declarative knowledge occurs in both, as well as complex learning involving refining and restructuring of knowledge. Existing knowledge can be tuned, either through specialization or generalization, and new structures can be built from old, either through generation of analogous structures, or through the combination and concatenation of old ones. This particular set of mechanisms of learning is not novel. *Flavell* [1972] described many of these ideas, and so did *Gagné* [1968], and more recently *Fischer* [1980]. The uniqueness of the two models that we have discussed derives from the specificity with which their mechanisms are described. Our hypothesis for the time being is that either one of these models is perfectly adequate to simulate development. In fact, it may well be useful to combine ideas from both of them. Schemas are very useful for describing general units of behavior, like the sets in *Fischer's* [1980] theory, and they can, in turn, be described as groups of production rules when a more specific analysis of structure is needed.

Interpretations from a Learning Framework

One of the difficulties in modelling development or any other important psychological change, such as the acquisition of expertise, is that there is an incredible amount of experience and knowledge involved in the change. Thus, a true developmental learning model, one which has the ability to acquire adult intelligence, is improbable if not impossible. However, if the same small set of processes is at work regardless of the domain or stage of development, a basic understanding is possible. *Keil* [1981] has pointed out that some set of cognitive constraints, constraints inherent in learning processes rather than in the knowledge to be learned, seems to be necessary in order to explain the efficiency with which we learn. How would children learn rules for generating language, for instance, if any possible generalization of what they hear is equally possible? Perhaps a small set of learning processes, such as those we have described, operating on a basic set of knowledge structures, provides these constraints. These processes will build up structures which are immensely complex, but which are based on a few principles.

If this is true, then models of limited domains and limited changes can illuminate the wider course of development. Smaller domains of increasing competence have been modelled successfully by changes in the knowledge structures available to the child. We have previously mentioned the success of the rule assessment method [*Siegler*, 1981], for instance. As another example, *Riley* et al. [in press] have modelled the knowledge required to perform various kinds of simple algebra word problems. They identified schemas which guide the representation of three basic problem types: change, compare and combine. *Riley* et al. [in press] were able to explain the performance of each child by showing that older children tend to have more accurate and complete versions of each schema. Although their model and the rule assessment models cannot explain how these different conceptual structures are acquired, it is clear that the components of the learning mechanisms proposed in the ASN and ACT models can eventually accommodate these transitions. With these thoughts in mind, we can now consider how a learning theory can explain the phenomena we described in the first section, as well as some others.

Stages and Decalage

One interesting consequence of a general learning model is that stages and decalage are really manifestations of a few underlying assumptions, but at different levels. In general, we assume: (1) only a small amount of new knowledge can be learned at any one time; (2) this new knowledge must be interpreted by and stored in existing knowledge structures; (3) new structures, when they are needed, are created from old ones, and (4) knowledge tends to be specific to the context in which it was learned.

Thus, when a child is learning a new skill, certain prerequisite knowledge and knowledge structures must already exist. If they do not, they must be learned first. For instance, *Siegler* [1981] showed that for a child to progress from a balance scale rule involving only weight to one involving distance as well, it is necessary for the child to learn to encode distance first, before using it in a rule. This idea was essentially proposed by *Gagné* [1968]. Specifically, he theorized that learning is necessarily hierarchical in nature; that in order to learn a concept such as conservation of liquid, a series of component pieces of knowledge must be learned first. These components may be procedural rules or declarative concepts, and each one, in turn, requires the existence of other subsidiary pieces of knowledge. For example, in order for conservation of liquid to be mastered, he proposed that a rule stating that the volume of liquid is determined jointly by its length, width and height (in a rectangular container) is necessary. He further proposed that a necessary preliminary to learning this rule is to learn three rules stating that if one dimension is held constant, changing one other dimension results in a compensatory change in the remaining one. The actual rules he proposed may not be correct; however, his proposal shows graphically how the content of the knowledge itself and the limits on how much can be learned at any one time create a stage-like learning sequence. This view is rather similar to *Fischer's* [1980] notion of *intercoordination* of sets and *Rumelhart and Norman's* [1976] *schema induction* in which a new schema is created from two or more older ones, as well as to *Flavell's* [1972] *hierarchical integration*. It can also be related to *composition* in *Anderson's* [1981] terms. The component rules would be learned individually and used separately at first, until, with practice they were 'added up' to form the conservation rule. Thus, stages may result from the interaction of constraints inherent in the knowledge being learned and in how knowledge is acquired and structured.

To understand decalage, it is necessary to assume that when conservation of X is learned, what is learned is a knowledge structure specific to X, not a general concept. That is not to say that there might not be a general concept of conservation which can be learned; however, such a concept is very complex due to the number of things which are conserved and the

number of transformations under which they are conserved. Some decalage may result because each new task really requires a more complex knowledge structure. For instance, conservation of weight can not be mastered if conservation of substance has not already been mastered, and the weight structure may be built up from the substance structure. Other kinds of decalage may result because subsidiary knowledge can not be mastered. Conservation of volume is typically not mastered until age 10 or 11, and it obviously cannot be mastered if the concept of volume itself is not understood. A study of volume alone might well show a sequence of preliminary knowledge states, taking some years to master, very similar to other sequences which have been noted.

Since learning is a process shared by adults, phenomena analogous to stages and decalage should be observable at all ages. We have already noted that there are observations in the adult literature which are analogous to decalage, such as lack of transfer on problem isomorphs. Since we see stages as simply a necessary step-by-step learning process, they too are evident in adults. An adult can no more learn a piece of knowledge without its prerequisites than can a child. However, an adult knowledge base is far more elaborate, and the kinds of new concepts and skills adults acquire are more complex than the simple sort of tasks *Piaget* pioneered. Thus, pinpointing the knowledge required to learn an adult skill is far more difficult and so is determining whether an adult has some or all of that knowledge. Also, for a more complex skill, there are undoubtedly many more possible sequences of knowledge states leading to the same result, making determination of the existence of any such sequences all the more difficult. In fact, a number of authors have pointed out that multiple pathways are probably available to developing children as well [*Fischer*, 1980; *Longeot*, cited in *Vuyk*, 1981]. Thus, when considering the invariant sequences thought to be characteristic of development it is important to remember that the invariance may be on a more abstract level than that of the actual chain of knowledge states. The constraints inherent in knowledge and the characteristics of learning may only limit development to a series of possible sequences, not a single invariant one.

Levels of Understanding

An interesting finding, related to stages and decalage, is that of levels of understanding. *Piaget* noted that once a child has acquired a particular principle, a rapid broadening of understanding of related phenomena takes place. As *Fischer* [1980, p. 485] put it, 'As a child moves into a new level, he or she will show rapid change, but once the level has been attained, he or she will show slower change'. There are a number of factors which can be seen from a learning perspective to produce this sort of effect. Although we have noted that knowledge is specific to the context in which it is learned, that is not to say that it is absolutely specific. When a child learns conservation of substance, for instance, it is unlikely that what is learned is a piece of knowledge that relates only to a specific piece of clay deformed in a certain way by a certain person, etc. Although this knowledge is not so general that it represents an abstract understanding of 'conservation', it must have some generality or it would never apply after the first time it was learned. The schema slot or production rule condition element which matches the deformed substance, for instance, should at least cover all kinds of clay and probably more substances that are similar to clay. By the same token, there are many elements of the situation, such as the time of day and the particular location, which should not be incorporated into the knowledge structure at all. How all this happens is not clear; however, there is evidence that children naturally tend to generalize their experiences [*Nelson* et al., this volume]. At any rate, newly acquired skills and knowledge are automatically ready to be used in situations which are somehow similar to the one in which they were learned, thus, a rapid broadening of understanding can proceed from the acquisition of one new structure.

The phenomenon of rapid change can also occur when new structures are created from old ones. For instance, in ACT, if two productions are found to be potentially applicable in a certain situation and they have enough in common in terms of the structures of their conditions and actions, a new *generalization* of the two can be produced. Of course, if this new production captured only what was in the previous two and nothing more, it would produce identical behavior. This is not the case, however, because specific elements in the two productions are converted to more general ones in the new production. Two different constants might be replaced by a variable; two different variables might be replaced by a more general variable. Thus, the new production has a wider applicability than the two on which it is based, and although it may well need to be refined through *discrimination* to finally achieve the proper scope, it has the potential for allowing wider understanding. Schemas can also be generalized by widening the scope of individual slots. In addition, new schemas, based on old ones, can be created through *patterned generation*. When no schema can be found which successfully applies to a given situation, a new one

is created based upon one that partially fits. It is important, of course, that a schema exists which does partially fit. Another way of saying this is that the new situation is analogous to a familiar one.

Hence, once a new structure is formed, it allows a new level of understanding into many related problems. This happens both because the new structure itself has some generality built in and because a given structure can spawn many new related ones through various processes, which might be summed up as learning by analogy. At first, there is a relatively large area of related problems and situations to which the new structure or analogous ones can apply and there is a rapid burst of new understanding; however, as the area remaining is quickly reduced, the process slows down.

Although we propose that new levels of understanding result from learning, we should emphasize that this does not mean that they cannot appear abruptly. It may require months or years to accumulate enough new facts and subsidiary structures to allow the creation of an important new structure, but once all of that is available, the proper situational demands can cause it to be built very rapidly. For instance, *Norman* [1978] has proposed that sudden insight, the 'aha' phenomenon, occurs due to *restructuring* of existing knowledge. He points out that there need be no addition of knowledge at all during this process and it might occur due to the demands of a particular situation, as in a Socratic tutorial. A similar process seems to take place in the development of knowledge in general, such as scientific knowledge. For instance, it took 100 years from the time that the Academy of Experiments in Florence discovered that freezing and boiling take place instantaneously at a certain 'degree of heat' until *Black* was able to differentiate heat from temperature [*Carey*, in press]. During this time, the discovery of many facts about heat and temperature set the stage for *Black's* realization.

Memory Deficits

There are two basic ways in which learning can affect memory performance: the acquisition of memory strategies and the acquisition of new knowledge per se. There is considerable evidence that both of these effects are important. In the case of strategies, the most commonly used example is rehearsal of items in short-term memory. Young children may not use it at all, while somewhat older ones may use it idiosyncratically or only in certain situations. Numerous efforts have been made to teach rehearsal and, indeed, it can be taught, although a common finding is that subjects will still not use it spontaneously. The important point, however, is that this strategy and, no doubt others, can be learned and will improve memory performance. There is evidence that failure to use it spontaneously may be the result of the unavailability of necessary content knowledge [*Chi*, in press, a].

The effect of content knowledge on memory has already been noted. We have shown that it is possible to reverse adult and child memory performance if the material to be recalled is familiar to the child and unfamiliar to the adult [*Chi*, 1978]. As an extreme example, consider the task of remembering a series of words containing one which is completely unfamiliar. In the case of this word, only its constituent sounds are familiar. Thus, while the other words might each activate one node representing the internal concept for which that word stands, the unfamiliar one activates a series of nodes representing its constituent sounds, thereby occupying several slots in short-term memory. More generally, the more familiar material is, the greater is the number of links between concept nodes and the more likely it is that several items can be grouped under one node. Each group of items can be called a declarative schema or a chunk [*Simon*, 1974]. *Chase and Simon* [1973a, b], for instance, found that chess masters can remember fairly complicated chess positions very accurately with only a 5-second exposure because they group the pieces into a few chunks. Only the nodes representing the chunks need to be in short-term memory and when the individual pieces are to be recalled, each chunk can be retrieved from long-term memory and unpacked. In addition, the greater number of links to a node resulting from greater familiarity might simply make access to that node easier and faster [*Chi*, 1976], so that more items can be rehearsed or retrieved in the same amount of time. Thus, the contents and structure of declarative memory have a very strong impact on short-term memory ability.

An outstanding example of the effects of both knowledge and strategies on memory skills is *Chase and Ericsson's* [1981] subject S.F. He was able to learn, through heroic amounts of practice, to recall as many as about 80 random digits, presented verbally at one per second. *Chase and Ericsson* [1981] have shown convincingly that his performance was due to three basic components: (1) a large store of factual knowledge related to numbers, (2) a retrieval structure in long-term memory, and (3) very highly refined encoding and retrieval strategies. The factual knowledge base he used was basically an extensive knowledge of times for various track events of different lenghts. He used 11 different distances from half-mile to marathon and several categories of times for each, such as world records and his own bests. He developed a highly skillful procedure to group the incoming digits and relate them to this knowledge base. For instance, 356 might be encoded as his old coach's best time for the mile (3:56).

This idea is not as simple as it may seem, however, because of the rapid sequential presentation of the digits. At first, he had to collect a series of digits and decide how to group them and what kind of running time to relate them to. Such conscious decisions require short-term memory space and take time to complete and, so, interfere with subsequent digits. As time progressed, he developed a highly automatic discrimination procedure which operated on each digit as it came in, successively narrowing down the number of possibilities. He also developed a retrieval structure, which *Chase and Ericsson* [1981] characterize as a directly addressable hierarchical long-term memory structure, rather like a schema. It specified how the digits were to be grouped before presentation began, eliminating any need for concurrent grouping decisions, provided a structure into which the numbers could be stored directly without having to use short-term memory and provided the means for ordering the groups upon retrieval. Thus, each group of digits was linked to a semantic structure representing the running-time mnemonic as well as to the retrieval structure through the use of highly specialized and efficient memory procedures. Not only did this structure enable SF to recall as many as 80 digits in one trial, he was also able to recall nearly all of the digit groups from an entire 1-hour session by accessing them through his 11 running-time categories. Although this is a highly specialized example, it indicates very clearly how important knowledge, knowledge structures and procedural skills are in simple memory tasks, such as a span task.

Learning to Learn

The importance of the products of learning in memorization skills brings up a broader question: Are there fundamental invariant learning processes or can these processes themeselves undergo changes? Although, S.F. 'learned to learn' strings of digits, there is no reason to believe that his more fundamental processes, such as generalization and discrimination, underwent any changes. However, in systems such as ACT and ASN, since the learning processes themselves can be represented as rules or schemas, they can thereby be accessible to each other. Thus, as *Langley and Simon* [1981] point out, a discrimination rule might act upon a generalization rule to produce a new rule which makes more accurate generalizations, perhaps because it is specific to a particular domain. A large supply of new versions of learning rules could be created which would be better suited to current learning demands upon the developing child or adult and which would make learning faster and more efficient.

This proposal is clearly speculative and such changes may be nearly impossible to detect and/or unnecessary because changes in strategies and knowledge can have a profound impact on learning. Note that we are making a distinction between basic learning processes (such as generalization or composition) and learning strategies (such as elaboration or rehearsal). For instance, elaboration, analyzing incoming information in terms of existing knowledge, can be a conscious learning strategy. Although many do it automatically, and it is partially dependent upon the associative nature of memory, there is no reason it could not be taught and learned and improved through practice. Further, the contents and structure of the knowledge base determine how sucessfully particular information can be elaborated. To continue with our example from a previous section, a child who is just starting to learn about dinosaurs might only have one schema for encoding dinosaurs. For this novice child, learning about a meat-eating dinosaur with sharp teeth and a short neck probably requires the storage of the particular dinosaur name along with all this property information. For the expert child, however, who has many subschemas of different types of dinosaurs, the dinosaur's characteristics fit immediately into the slots of the meat-eater schema. The information that it is a meat eater is simply redundant confirmation that the right schema has been activated. This schema gives immediate access to examples of other meat-eaters, and the child can compare them with the new example in order to determine its discriminating characteristics and encode them. This example serves to point out how the declarative encoding process can vary as a function of existing knowledge structure. In this situation, because the expert child need not encode the basic features common to the meat-eater schema and has rapid access to other specific examples, a more sophisticated learning strategy may be employed. The novice child may also be able to compare examples and seek discriminating features but, in this case, the knowledge base does not allow it.

Thus, in a sense, learning to learn is definitely possible. However, it need not be basic learning processes that are learned or improved, but higher level strategies, and even in cases where similar strategies are available, the interaction of these strategies with the knowledge base is very important. A more complete knowledge base may allow more efficient learning without any differences at all in learning strategies and processes.

Acknowledgement

Preparation of this paper was supported in part by the Learning Research and Development Center, University of Pittsburgh, which is funded in part by the National Institute of Education (NIE). Further support was provided by the Spencer Foundation. The opinions expressed do not necessarily reflect the position or policy of the NIE or the Spencer Foundation and no official endorsement should be inferred. The authors would like to thank *Mary Beth Curtis* and *Robert Glaser* for their comments on earlier drafts of this paper.

References

Anderson, J.R.: Language, memory and thought (Erlbaum, Hillsdale 1976).

Anderson, J.R.: Acquisition of cognitive skill. Tech. Rep. 81-1 (Carnegie-Mellon University, Department of Psychology, Pittsburgh 1981).

Anderson, J.R.; Bower, G.H.: Human associative memory (Winston, Washington 1973).

Anderson, J.R.; Greeno, J.G.; Kline, P.J.; Neves, D.M.: Acquisition of problem-solving skill; in Anderson, Cognitive skills and their acquisition, pp. 191–230 (Erlbaum, Hillsdale 1981).

Bartlett, F.C.: Remembering: a study in experimental and social psychology (Cambridge University Press, Cambridge 1932).

Baylor, G.W.; Gascon, J.: An information processing theory of aspects of the development of weight seriation in children. Cognitive Psychol. *1:* 1–40 (1974).

Brown, A.L.: The development of memory: knowing, knowing about knowing, knowing how to know; in Reese, Advances in child development and behavior, vol. 10, pp. 104–152 (Academic Press, New York 1975).

Brown, A.L.: Learning and development: the problems of compatibility, access, and induction. Hum. Dev. *25:* 89–115 (1982).

Brown, A.L.; Campione, J.C.: Inducing flexible thinking: a problem of access; in Friedman, Das, O'Connor, Intelligence and learning, pp. 515–529 (Plenum Press, New York 1981).

Bruner, J.S.; Olver, R.R.; Greenfield, P.M.: Studies in cognitive growth (Wiley, New York 1966).

Carey, S.: Are children fundamentally different kinds of thinkers and learners than adults?; in Chipman, Segal, Glaser, Thinking and learning skills: current research and open questions, vol. 2 (Erlbaum, Hillsdale, in press).

Case, R.: Validation of a neo-Piagetian capacity construct. J. exp. Child Psychol. *14:* 287–302 (1972).

Chase, W.G.; Ericsson, K.A.: Skilled memory; in Anderson, Cognitive skills and their acquisition, pp. 141–189 (Erlbaum, Hillsdale 1981).

Chase, W.G.; Simon, H.A.: Perception in chess. Cognitive Psychol. *4:* 55–81 (1973a).

Chase, W.G.; Simon, H.A.: The mind's eye in chess; in Chase, Visual information processing (Academic Press, New York 1973b).

Chi, M.T.H.: Short-term memory limitations in children: capacity or processing deficits? Memory Cognition *4:* 559–572 (1976).

Chi, M.T.H.: Knowledge structures and memory development; in Siegler, Children's thinking: What develops?, pp. 73–96 (Erlbaum, Hillsdale 1978).

Chi, M.T.H.: Interactive roles of knowledge and strategies in the development of organized sorting and recall; in Chipman, Segal, Glaser, Thinking and learning skills: current research and open questions, vol. 2 (Erlbaum, Hillsdale, in press, a).

Chi, M.T.H.: Knowledge-derived categorization in young children; in Rogers, Sloboda, Acquisition of symbolic skills (Plenum Press, New York, in press, b).

Chi, M.T.H.; Koeske, R.D.: Network representation of a child's dinosaur knowledge. Devl. Psychol. *19:*29–39 (1983).

Childs, M.K.; Polich, J.M.: Developmental differences in mental rotation. J. exp. Child Psychol. *27:* 339–351 (1979).

Collins, A.; Loftus, E.F.: A spreading activation theory of semantic processing. Psychol. Rev. *82:* 407–428 (1975).

Collins, A.; Quillian, M.R.: Retrieval time from semantic memory. J. verb. Learn. verb. Behav. *8:* 240–247 (1969).

Cooper, L.A.; Shepard, R.N.: Chronometric studies of the rotation of mental images; in Chase, Visual information processing, pp. 75–176 (Academic Press, New York 1973).

Corsini, D.A.: Memory: interaction of stimulus and organismic factors. Hum. Dev. *14:* 227–235 (1971).

Dempster, F.M.: Memory span and short-term memory capacity: a developmental study. J. exp. Child Psychol. *26:* 419–431 (1978).

Ericsson, K.A.; Simon, H.A.: Verbal reports as data. Psychol. Rev. *87:* 215–253 (1980).

Fischer, K.W.: A theory of cognitive development: the control and construction of hierarchies of skills. Psychol. Rev. *87:* 477–531 (1980).

Flavell, J.H.: The developmental psychology of Jean Piaget (Van Nostrand, Princeton 1963).

Flavell, J.H.: Development studies of mediated memory; in Reese, Lipsitt, Advances in child development and behavior, vol. 5, pp. 182–211 (Academic Press, New York 1970).

Flavell, J.H.: First discussant's comments: What is memory development the development of? Hum. Dev. *14:* 272–278 (1971).

Flavell, J.H.: An analysis of cognitive developmental sequences. Genet. Psychol. Monogr. *86:* 279–350 (1972).

Gagné, R.M.: Contributions of learning to human development. Psychol. Rev. *75:* 177–191 (1968).

Gallagher, J.D.; Thomas, J.R.: Effects of varying post-KR intervals upon children's motor performance. J. Motor Behav. *12:* 41–56 (1980).

Gelman, R.: Conservation acquisition: a problem of learning to attend to relevant attributes. J. exp. Child Psychol. *7:* 167–187 (1969).

Gelman, R.: Cognitive development. A. Rev. Psychol. *29:* 297–332 (1978).

Gesell, A.: Infancy and human growth (Macmillan, New York 1928).

Gesell, A.; Thompson, H.: Learning and growth in identical twin infants. Genet. Psychol. Monogr. *6:* 1-24 (1929).

Halford, G.S.; Macdonald, C.: Children's pattern construction as a function of age and complexity. Child Dev. *48:* 1096–1100 (1977).

Halford, G.S.; Wilson, W.H.: A category theory approach to cognitive development. Cognitive Psychol. *12:* 356–411 (1980).

Hayes, J.R.; Simon, H.A.: Psychological differences among problem isomorphs; in Castellan, Pisoni, Potts, Cognitive theory, vol. 2, pp. 21–41 (Erlbaum, Hillsdale 1977).

Johnson-Laird, P.N.; Wason, P.C.: A theoretical analysis of insight into a reasoning task. Cognitive Psychol. *1:* 134–148 (1970).

Kail, R.V.; Hagen, J.W.: Perspectives on the development of memory and cognition (Erlbaum, Hillsdale 1977).

Kail, R.V.; Pellegrino, J.; Carter, P.: Developmental changes in mental rotation. J. exp. Child Psychol. *29:* 102–116 (1980).

Keil, F.C.: Constraints on knowledge and cognitive development. Psychol. Rev. *88:* 197–227 (1981).

Klahr, D.; Wallace, J.G.: Class inclusion processes; in Farnham-Diggory, Information processing in children, pp. 143–172 (Academic Press, New York 1972).

Kosslyn, S.D.: The representational-development hypothesis; in Ornstein, Memory development in children, pp. 157–189 (Erlbaum, Hillsdale 1978).

Langley, P.; Simon, H.A.: The central role of learning in cognition; in Anderson, Cognitive skills and their acquisition, pp. 361–380 (Erlbaum, Hillsdale 1981).

Lawler, R.W.: The progressive construction of mind. Cognitive Sci. *5:* 1–30 (1981).

Liben, L.S.: The developmental study of children's memory; in Field, Huston, Quay, Troll, Finly, Review of human development (Wiley, New York, in press).

Lindberg, M.A.: Is knowledge base development a necessary and sufficient condition for memory development? J. exp. Child Psychol. *30:* 401–410 (1980).

Mandler, G.: The construction and limitation of consciousness. Proc. Int. Symp. on Perspectives on Psychological Experimentation, Bad Homburg 1981.

Mandler, J.M.: Representation; in Flavell, Markman, Cognitive development, vol. 2 of Mussen, Carmichael's manual of child psychology (Wiley, New York, in press).

Marmor, G.S.: Development of kinetic images: When does the child first represent movement in mental images? Cognitive Psychol. *7:* 548–559 (1975).

McLaughlin, G.H.: Psycho-logic: a possible alternative to Piaget's formulation. Br. J. educ. Psychol. *33:* 61–67 (1963).

Miller, G.A.: The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol. Rev. *63:* 81–97 (1956).

Minsky, M.A.: A framework for representing knowledge; in Winston, The psychology of computer vision (McGraw-Hill, New York 1975).

Nelson, K.: Semantic development and the development of semantic memory; in Nelson, Children's language, vol. 1, pp. 39–80 (Gardner Press, New York 1978).

Neves, D.M.; Anderson, J.R.: Knowledge compilation: mechanisms for the automatization of cognitive skills; in Anderson, Cognitive skills and their acquisition, pp. 57–84 (Erlbaum, Hillsdale 1981).

Newell, A.: Production systems: models of control structures; in Chase, Visual information processing, pp. 463–526 (Academic Press, New York 1973).

Newell, A.; Simon, H.A.: Human problem solving (Prentice Hall, Englewood Cliffs 1972).

Norman, D.A.: Notes toward a theory of complex learning; in Lesgold, Pellegrino, Fokkema, Glaser, Cognitive psychology and instruction (Plenum, New York 1978).

Norman, D.A.; Rumelhart, D.E.: Explorations in cognition (Freeman, San Francisco 1975).

Olson, G.M.: Developmental changes in memory and the acquisition of language; in Moore,

Cognitive development and the acquisition of language, pp. 145–157 (Academic Press, New York 1973).
Ornstein, P.A.: Memory development in children (Erlbaum, Hillsdale 1978).
Ornstein, P.A.; Naus, M.J.: Rehearsal processes in children's memory; in Ornstein, Memory development in children, pp. 69–99 (Erlbaum, Hillsdale 1978).
Paivio, A.: Imagery and verbal processes (Holt, New York 1971).
Pascual-Leone, J.: A mathematical model for the transition rule in Piaget's developmental stages. Acta psychol. *32:* 301–345 (1970).
Piaget, J.: Judgment and reasoning in the child (Harcourt, Brace & World, New York 1928).
Piaget, J.: The theory of stages in cognitive development; in Green, Ford, Flamer, Measurement and Piaget, pp. 1–11 (McGraw-Hill, New York 1971).
Piaget, J.: Intellectual evolution from adolescence to adulthood. Hum. Dev. *15:* 1–12 (1972).
Quillian, M.R.: Semantic memory (Bolt, Beranek & Newman, Cambridge 1966).
Riley, M.S.; Greeno, J.G.; Heller, J.I.: Development of children's problem-solving ability in arithmetic; in Ginsberg, The development of mathematical thinking (Academic Press, New York, in press).
Robertson, M.A.: Longitudinal evidence for developmental stages in the forceful overarm throw. J. Hum. Movmnt Stud. *4:* 167–175 (1978).
Rosch, E.; Lloyd, B.B.: Cognition and categorization (Erlbaum, Hillsdale 1978).
Rozin, P.: The evolution of intelligence and access to the cognitive unconscious; in Sprague, Epstein, Progress in psychobiology and physiological psychology, vol. 6, pp. 245–280 (Academic Press, New York 1976).
Rumelhart, D.E.; Lindsay, P.; Norman, D.A.: A process model for long-term memory; in Tulving, Donaldson, Organization of memory (Academic Press, New York, 1972).
Rumelhart, D.E.; Norman, D.A.: Accretion; tuning and restructuring: three modes of learning, Tech. Rep. CHIP 63 (University of California, San Diego Center for Human Information Processing, La Jolla 1976).
Rumelhart, D.E.; Norman, D.A.: Analogical processes in learning; in Anderson, Cognitive skills and their acquisition, pp. 335–359 (Erlbaum, Hillsdale 1981).
Rumelhart, D.E.; Norman, D.A.: Simulating a skilled typist: a study of cognitive-motor performance. Cognitive Sci. *6:* 1–36 (1982).
Rumelhart, D.E.; Ortony, A.: The representation of knowledge in memory; in Anderson, Spiro, Montague, Schooling and the acquisition of knowledge (Erlbaum, Hillsdale 1977).
Scardamalia, M.: Information processing capacity and the problem of horizontal decalage: a demonstration using combinatorial reasoning tasks. Child Dev. *48:* 28–37 (1977).
Schank, R.C.; Abelson, R.P.: Scripts, plans, goals and understanding: an inquiry into human knowledge structures (Erlbaum, Hillsdale 1977).
Siegler, R.S.: Three aspects of cognitive development. Cognitive Psychol. *8:* 481–520 (1976).
Siegler, R.S.: Developmental sequences within and between concepts. Soc. Res. Child Dev. Monogr. *46:* (1981).
Siegler, R.S.; Robinson, M.: The development of numerical understandings; in Reese, Lipsett, Advances in child development and behavior, vol. 16 (Academic Press, New York 1981).
Simon, H.A.: On the development of the processor; in Farnham-Diggory, Information processing in children, pp. 3–22 (Academic Press, New York 1972).
Simon, H.A.: How big is a chunk? Science *183:* 482–488 (1974).

Vuyk, R.: Overview and critique of Piaget's genetic epistemology (Academic Press, London 1981).
Waterman, D.A.: Adaptive production systems. Proc. 4th Int. Joint Conf. on Artificial Intelligence, pp. 296–303 (1975).
Young, R.M.: Strategies and the structure of a cognitive skill; in Underwood, Strategies of information processing, pp. 35–45 (Academic Press, London 1978).