logical-deduction task, in accord with the view that this task required skills that were not part of the training program. In sum, high-IQ children’s superior knowledge-acquisition components contributed to their superior performance on the insight problems.

The third part of Sternberg’s theory involves metacomponents, components used to construct strategies. Metacomponents govern the use of the other components. They also are responsible for most aspects of developmental change. As Sternberg (1984) commented “There can be no doubt that in the present conceptual scheme, that metacomponents form the major basis for the development of intelligence” (p. 172).

The importance of metacomponents is evident in people’s transfer of knowledge from one context to another. Older children and people with greater expertise are generally better able to apply their knowledge to new problems than are younger people and people with less expertise (Campione & Brown, 1984; Gentner, Ratterman, Markman, & Kотовsky, 1995; Staszewski, 1988). Knowledge is especially important; 10-year-olds who are expert at chess more successfully solve novel chess problems than adults with little knowledge of chess but whose general memory capacities are higher (Chi, 1978). However, within a given level of knowledge, people with higher IQs generally can apply existing knowledge to acquire new knowledge more rapidly (e.g., Johnsrød & Mervis, 1994).

How should Sternberg’s theory be evaluated? Two weaknesses can be noted. One is that the theory summarizes more than it predicts. It is not clear what types of evidence would be inconsistent with the approach. Another involves the role of metacomponents in the organization of the system. These are crucial parts of the overall theory, but their workings remain somewhat mysterious. On the other hand, the theory is exceptional in the breadth of phenomena and of populations to which it has proven applicable. It encompasses a large number of intuitively important aspects of development and organizes them in an easy-to-grasp way. It provides a plausible outline of how a strategy-construction mechanism would operate. In short, it constitutes a useful framework within which to view development.

Production-System Theories

Perhaps the most difficult challenge for theories of cognitive development has been to explain how development occurs. Piaget and many others have tried to generate such explanations, but they have not been entirely successful. Consider the following evaluation:

For 40 years now we have had assimilation and accommodation, the mysterious and shadowy forces of equilibration, the Batman and Robin of the developmental processes. What are they? How do they do their thing? Why is it after all this time, we know no more about them than when they first sprang on the scene? What we
need is a way to get beyond vague verbal statements of the nature of the development. (Kluft, 1982, p. 80)

One promising effort to provide more precise and satisfying explanations of change has been to model development through production systems. These are a class of computer-simulation languages that have proven useful for modeling cognitive development. Each production is a kind of if-then rule that indicates what the system would do in a particular situation. Together, the productions indicate what the system would do under a wide range of circumstances. The key properties of production systems are:

1. The basic organization consists of two interacting structures: a production memory, which is the system's enduring knowledge, and a working memory, which is the system's representation of the current situation.
2. The production memory includes a large number of specific productions, each of which includes a condition side and an action side.
3. The condition side of each production specifies the circumstances under which the production is applicable. The action side specifies the actions that are taken when these conditions are met. Such actions include both activities in the external world and manipulations of symbols in working memory.
4. The contents of working memory are constantly changing, because they reflect constantly changing situations. Information enters working memory both through perception of events in the external world and through taking the actions indicated by the action side of productions.
5. Thinking occurs through a cycle of (a) information being present in working memory, (b) the information matching the condition side of one or more productions, (c) this match resulting in the actions on the action side of those productions being taken, (d) the actions placing new information in working memory, thus starting the cycle anew.
6. Learning occurs through a process of self-modification, in which new productions are created and existing productions modified as a result of previous experience.

The basic organization of production systems is diagrammed in Figure 3.5.

An example of a simple production system that produces correct performance on Piaget's number conservation problem is shown in Table 3.3. The bottom part of the table indicates the sequence of working-memory states that the system produces while solving the problem. The particular production system always searches downward from the top of the list of productions until it finds a production whose condition side is matched by the contents of working memory. That production then fires, and the search begins anew from the top of the list.

In the experimental situation to which the Table 3.3 production system applies, the child has been shown two rows of objects, has been told that they have the same number of objects, has seen the objects in one of the rows spread out, and has been asked whether the two rows now have the same number of objects. This information is represented in the initial contents of working memory in the bottom part of Table 3.3. The initial state of working memory matches the condition side of P1 (Production 1), which therefore fires, putting into working
memory a goal of stating the numerical relation between the rows. When the system starts from the top again, the contents of working memory do not match P1 (because its second condition is not matched) nor P2 (because its second condition is not matched.) However, the contents of working memory do match the condition side of P3, which therefore fires. This places in working memory the information that the rows have the same number of objects. With this information, P2 can fire and the system states the correct answer.

David Klahr is probably the most prominent advocate of production systems as a tool for explaining how development occurs. The key developmental mechanism in Klahr’s theory is generalization. Number conservation provides a convenient context for explaining how his theory works.

Klahr and Wallace (1976) divided the process of generalization into three components: the time line, regularity detection, and redundancy elimination. The time line contains the data on which generalizations are based. It is a record of all the situations the system has ever encountered, the responses produced in those situations, the outcomes of the actions, and the new situations that arose. Table 3.4 illustrates the type of information that might be included in the time line’s record of a single event. A child saw a group of cookies and noticed that
TABLE 3.3  A simple production system for number conservation*

P1: if you are asked about the numerical relation between two collections and you do not have a goal of stating the relation, then set a goal of stating the relation.

P2: if you have a goal of stating the numerical relation between two collections and you know the relation, then state the relation.

P3: if you have a goal of stating the numerical relation between two collections, and the collections had the same number of objects before a transformation and the transformation did not involve adding or subtracting objects, then the rows still have the same number of objects.

Initial Working Memory (WM1): Rows had same number of objects before, one row then was spread, nothing added or subtracted, question is whether rows have same number of objects now.

P1 fires.

WM2: Goal is to state whether rows have same number of objects now, rows had same number of objects before, one row then was spread, nothing added or subtracted, question is whether rows have same number of objects now.

P3 fires.

WM3: Goal is to state whether rows have same number of objects now, rows have the same number of objects, rows had same number of objects before, one row then was spread, nothing added or subtracted, question is whether rows have same number of objects now.

P2 fires.

System answers: “The rows have the same number of objects.”

Source *Adapted from F贾V & Wallace, 1976.

there were three. This realization was made possible by subitizing (a process by which both children and adults can rapidly perceive the number of objects in sets ranging from one to four objects). Next, the child transformed the spatial position of the cookies by picking them up in his hand. Finally, the child again subitized the collection of cookies and found that there still were three.

Such detailed records of situations, responses, and outcomes might at first seem unnecessary. Why remember so much about each experience? In fact, the information could be invaluable. In many situations, children cannot know beforehand what will turn out to be relevant. If they retain detailed information that may or may not be relevant, they later may be able to draw unanticipated generalizations. If they retain only what they know to be relevant, however they will miss much relevant information.

Is it realistic to think that children have a memory record similar to a time line? Observing the level of detail with which they remember certain information suggests that it is. Almost all parents have anecdotes to this effect. One of
TABLE 3.4. A Portion of a Child's Time Line

<table>
<thead>
<tr>
<th>(PREVIOUS PROCESSING EPISODES)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
87456. Cookies on table.
87457. I subtitized.
87458. There were three.
87459. I heard a bird.
87460. I picked up the cookies.
87461. I subtitized the cookies.
87462. There were three again.

mine concerns a vacation on which my wife, our almost-2-year-old son, and I were staying in a motel. We wanted to go to dinner but could not find the room key. After 10 minutes of searching, I finally listened to my son long enough to understand what he was saying: "Under phone." As soon as I understood, I knew he was right. I had put it there (for reasons I no longer remember). It seems likely that if he remembered this relatively inconsequential detail, he probably was remembering many other details as well. Hasher and Zacks's (1984) ideas about automatic processing of frequency information and of several other aspects of experience, such as spatial locations and time of occurrence, suggest the types of content that might be entered into the time line. Thus, Klahr and Wallace's contention that children retain a detailed ledger of their experiences seems quite plausible.

The second key process, regularity detection, operates on the contents of the time line to produce generalizations about experience. This is accomplished by the system's noting places in the time line where many features are similar and where the same outcome occurs despite variations in one or more features. In number conservation, regularity detection could produce at least three types of generalizations. One would involve generalizing over different objects. Regardless of whether two checkers, two coins, two dolls, or two cookies were spread, there would still be two objects. Children also could generalize over equivalent transformations. Spreading, compressing, piling up, and putting in a circle all preserve the initial number of objects.

The third process in Klahr and Wallace's model, redundancy elimination, accomplishes a different type of generalization. It improves efficiency by identifying processing steps that are unnecessary, thus reaching the generalization that a shorter sequence can achieve the same goal. In the number-conservation example, children would eventually note that it is unnecessary to subitize after picking up the cookies. Since there were three cookies before and since picking
up objects never affects how many there are, the number still must be the same. Klahr and Wallace hypothesized that the information-processing system eliminates redundancy by examining procedures within the time line and checking if the same outcome always occurs even if one or more steps are deleted. If so, the simpler procedure is substituted for the more complex one.

When does the information-processing system have time to detect regularities and to eliminate redundancies? Klahr and Wallace (1976) advanced one intriguing possibility: Perhaps children do it in their sleep. Other possibilities are that moments of quiet play, relaxation, or daydreaming are when children accomplish these functions.

Klahr and Wallace’s approach, unlike stage theories, implies that different children develop skills in different orders. In the cognitive system’s attempts at self-modification, there is no reason why one type of regularity always should be detected before another type. Children learning about number conservation either could first detect that it does not matter if the rows of objects contain cookies or checkers or could first detect that it does not matter if the row of cookies is shortened or lengthened. Thus, there is less of a lock-step feel to the model than to stage approaches.

Another implication of Klahr and Wallace’s theory relates to the idea of encoding. The way in which information is encoded in the time line shapes the learning that can later occur. Suppose, for example, that in a liquid-quantity-conservation experiment, a child encodes only the heights of the water in the glasses. Such a child would not be able to detect the regular relation between increments in the height of water and decrements in its cross-sectional area. The information about cross-sectional area simply would not be available in the time line.

Klahr has been in the forefront of investigators arguing for greater use of computer simulation as a tool for modeling development. He has noted that such simulations allow more explicit and precise models of how development occurs than would otherwise be possible (Klahr, 1989; 1992). Consistent with this stance, Simon and Klahr (1995) formulated a self-modifying production system that illustrated how children could come to understand conservation. At the outset, the model could not solve the number conservation problems it was presented. Through experience trying to solve them, it figured out how to do so. Of special interest, Simon and Klahr generated two versions of the model: one corresponding to 3-year-olds and one to 4-year-olds. Both models were able to learn when given relatively extensive experience with the problems, but only the model of 4-year-olds learned from limited experience with them. These data corresponded to the results obtained with real 3- and 4-year-olds who had been presented with these experiences by Gelman (1982).

The models of the younger and older children suggested hypotheses concerning why 3- and 4-year-olds showed the patterns of learning that they did. Both models contained learning mechanisms that allowed them to learn from the more extensive experience. However, two differences between them resulted
in the model of 4-year-olds, but not the model of 3-year-olds, learning from the limited experience. The model of 4-year-olds more clearly remembered the relationship between the sets before the transformation, and it was more likely to check whether the differences between the lengths of the rows after the transformation corresponded to a difference in numbers of objects. These differences in the models were consistent with what is known generally about 3- and 4-year-olds. The 4-year-olds are more likely to use counting to check whether their perceptions regarding numbers of objects are correct (Sophian, 1987) and also usually remember more about past states (Schneider & Bjorklund, in press). Thus, the differences between the models of 3- and 4-year-olds were both consistent with past observations of these age groups and suggested hypotheses regarding why the two age groups might learn as they did in this particular context.

Not everyone shares Klahr’s enthusiasm for computer-simulation models, though. Critics note that people are not computers and that unlike computers, people develop. This leads them to the conclusion that development cannot be modeled appropriately on a computer (Belin, 1983; Liben, 1987).

As Klahr (1989) pointed out, however, ideas about development are embodied in the computer program, not the computer on which the program runs. The computer is simply the device used to test whether these ideas account for the known phenomena. To illustrate the point, Klahr noted that computer simulations of cognitive development do not imply that children are computers any more than computer simulations of hurricanes imply that the atmosphere is a computer.

Several limitations of Klahr’s theory should be mentioned. Although he has often proclaimed the virtues of self-modifying production systems, neither he nor other investigators interested in children’s thinking has yet written many of them. In addition, such self-modifying production systems thus far has been more useful for explaining previous findings than for generating new ones. On the other hand, these shortcomings do not detract from the potential of self-modifying production systems as models of development. In addition, Klahr and Wallace’s explanation of generalization in terms of the time line, regularity detection, and redundancy elimination is more precise and explicit than almost all other mechanisms of cognitive development that have been proposed. These are important virtues and may foreshadow additional breakthroughs.

Connectionist Theories

One especially “hot” approach to thinking about cognitive development (and to thinking about cognition in general) is connectionism. Like production systems, connectionist theories are computer simulations of how thinking occurs. Much of the reason for the rapidly-growing popularity of connectionist models is their general resemblance to the workings of the brain. This makes the approach a promising candidate for modeling how thinking is achieved within the brain. Connectionist models have several key characteristics (Plunkett, 1996):
1. They are made up of large numbers of simple processing units, akin to neurons in the brain.

2. The processing units are organized into two or more hierarchically-organized layers (Figure 3.6). Typically, these include an input layer, whose processing units encode the initial representation of the situation; one or more hidden layers, whose units combine information from the input units and an output layer, whose units generate the system’s response to the situation.

3. The individual processing units are connected to other processing units in different layers (and sometimes within the same layer as well). The strength of the connections varies with the system’s experience and is crucial in determining the processing that is done.

4. As in the brain, a given processing unit fires when the amount of activation it receives from all of the other processing units that are connected to it exceeds a threshold. The amount of activation that a unit receives from each unit connected to it is determined by the degree of activation of the processing unit that is sending the activation and the strength of the connection between the units.

5. As in the brain, the activity of the many simple processing units occurs in parallel (simultaneously).

6. Knowledge is represented through the strengths of connections among all of the units in the system. There is no single location that corresponds to a particular piece of knowledge; rather, the knowledge is distributed over all of the units and their interconnections. Because of this, and because processing occurs over many units in parallel, these systems are often known as parallel distributed processing (PDP) systems.

7. Learning occurs through the system receiving input, generating a response, observing the discrepancy between it and the correct answer, and adjusting the strengths of connections among the processing units in ways that would have led to a better answer. The adjustments include strengthening some connections and weakening others. Through this process, the system implicitly learns the rules underlying correct responses to the problem, although there is no single place in which the rule is represented.

8. Generalization of the system’s knowledge is based on similarity of new situations to ones the system has encountered previously. When the same types of implicit rules apply to new problems, connectionist systems are very effective in generalizing previous experience to them.

A number of researchers have advocated connectionist models of development: McClelland (1995); Shultz, Schmidt, Buckingham, and Maeschal (1995); Plunkett (1996); and Marchman (1992), among them. A particularly impressive connectionist model, and one that illustrates the strength of the approach for modeling development, is that of MacWhinney, Leinbach, Taraban, and McDonald (1989).

The model of MacWhinney et al. (Figure 3.6) depicted German children’s learning of their language’s system of definite articles. These definite articles are the multiple terms that in German serve the function that the single word the serves in English. The task was of interest precisely because the German article system is so difficult. Which article should be used to modify a given noun depends on the gender of the noun being modified (masculine, feminine, or neuter), its number (singular or plural), and its role within the sentence (subject,
FIGURE 3.6. The connectionist model of MacWhinney et al. (1989) of how children learn the German system of articles. Note that at the top (input) level, the model encodes five semantic features of the noun (corresponding to the "S" at the level just below the top on the extreme left), presence or absence of 11 phonological features at as many as 13 locations in the word (starting at the level just below the top), and 17 explicit case cues, which indicate the function of the noun within the sentence. These input-level units transfer activation to hidden units at the next two levels below, and eventually to the six output units, which correspond to the six articles that accompany nouns in German. The article that goes with the most activated output unit is advanced as the response (after MacWhinney et al., 1989). Copyright © 1989 by Academic Press, Inc.
possessor, direct object, or indirect object). To make matters worse, assignment of nouns to gender categories is often nonintuitive. For example, the word for fork is feminine, the word for spoon is masculine, and the word for knife is neuter. The relations are so complex that they seem almost impossible to learn. However, MacWhinney et al. built a connectionist model that showed how children could learn them.

The model of MacWhinney et al., like most connectionist models, involves an input layer, several hidden layers, and an output layer (Figure 3.6). Each of these layers contains a number of discrete units. For example, in the model, the 35 units within the input layer represent features of the particular noun that the article modifies. Each of the hidden layers includes units that represent combinations of these input-level features. The six output units represent the six articles in German that correspond to the in English.

As noted above, a central feature of such connectionist models is the very large number of connections among processing units. In the model of MacWhinney et al., each input-layer unit is connected to first-level hidden units; each first-level hidden unit is connected to second-level hidden units; and each second-level hidden unit is connected to each of the six output units. Learning occurs through a cycle of the system (1) receiving initial input (in this case, a noun in a certain context); (2) projecting on the basis of the strengths of its various connections (which reflect past experience) what output to produce; (3) advancing that response; and (4) adjusting the strengths of connections between units so that connections that suggested the correct answer are strengthened and connections that suggested the wrong answer are weakened.

MacWhinney et al. tested this system’s ability to master the German article system by repeatedly presenting it with 98 common German nouns. The simulation needed to choose which article to use with each noun in the particular context—that is, in the context of wanting to express a particular meaning with particular words. After it did this, the correct answer was presented, and the simulation adjusted connection strengths so as to optimize its accuracy in the future.

Following experience with this training set, the MacWhinney et al. simulation chose the correct article for more than 90 percent of the nouns in the original set. This could not be attributed simply to rote learning of which article accompanied each noun. When the simulation was presented a previously encountered noun in a novel context, it chose the correct article on more than 90 percent of trials, despite the noun’s often taking a different article in the new context than it had in the previous ones. The simulation also proved able to generalize to novel nouns; even when it had never encountered the particular term, it could use the term’s sound and meaning to make educated guesses as to what article would accompany it.

The simulation’s learning paralleled children’s learning in a number of ways. Early in the learning process, the simulation, like children whose first language is German, tended to overuse the articles that accompany feminine nouns. The reason appeared to be that this form of the articles is used most often within
the language. Further, the same article–noun combinations that are the most difficult for German children to learn were the most difficult for the simulation to learn as well. The particular errors made by the simulation also resembled those of children.

Connectionist models have successfully depicted a number of other developmental acquisitions as well. These include object permanence (Munakata, McClelland, Johnson, & Siegler, in press), understanding of balance scales, time–speed–distance problems, and causal reasoning (Shultz et al., 1995), early reading acquisition (Plaut, McClelland, Seidenberg, & Patterson, 1995), second language learning (MacWhinney, 1996), and acquisition of word meanings and grammatical understanding (Elman, 1992; MacWhinney & Chang, 1995; Marchman, 1992; Plunkett & Sinha, 1992).

As with all theories, connectionist approaches are open to criticism. One frequent criticism is that their claim to be “brain style cognition” is overstated. Nothing within them corresponds to the chemical activity that is crucial to brain functioning, and the functioning of their simple processing units bears only an abstract similarity to the functioning of neurons. Another limitation is that their learning is extremely slow and requires many more exposures than do human beings. They do not show the kind of sudden insight that people sometimes do (Rajimekers, Koten, & Molenaar, 1996). A third limit, related to the second one, is that they do not learn the symbolic rules, such as mathematical formulas, that people do, and may not be able to learn certain aspects of grammar (Pinker & Prince, 1988). On the other hand, connectionist models have proven useful for modeling the many developments that do not depend on acquisition of explicit rules. Although the operation of such systems clearly differs from that of the brain, it more closely resembles it than do other computer simulation approaches. Connectionist models have proven especially useful for modeling domains such as perception and language, in which numerous, partially valid sources of information must be integrated to produce successful performance. Given these advantages, it is not surprising that the popularity of connectionist models of development is growing rapidly.

Theories of Cognitive Evolution

One of the most profound intellectual contributions of all time is Darwin’s theory of evolution. Within evolutionary theory, competition among species is a basic aspect of existence. Species originate and change through two main processes: variation and selection. Genetic combination and mutation produce variation; survival of offspring is the basis of selection. Together, these processes have produced our planet’s ever-changing mosaic of living things.

As in the biological context, competition seems to be a basic feature of cognition. Rather than species competing, however, the competitions are among ideas. The main challenges for evolutionary theories of cognitive development are to describe the competing entities within the human cognitive system, to
describe how the competition among these entities leads to adaptive outcomes, and to identify the mechanisms that produce cognitive variation and selection. A number of recent models of cognitive development are based on analogies between the functions that must be accomplished to produce evolutionary and developmental change (Changeux & Dehaene, 1989; Cziko, 1995; Edelman, 1987; Johnson & Gilmore, 1996). However, since I'm the one writing the book, I will use my own overlapping waves approach to illustrate the way in which the analogy to biological evolution can contribute to understanding of development.

The basic assumptions of this approach are that at any one time, children have a variety of ways of thinking about most topics; that these varied ways of thinking compete with each other for use; and that the more advanced ways of thinking gradually become increasingly prevalent. These assumptions are illustrated in Figure 3.7. At any given time, several ways of thinking (the strategies in the figure) are present in a child’s thinking. (Strategies are procedures aimed at meeting particular goals.) These strategies compete with each other, and with experience, some become more frequent, some become less frequent, and some first become more frequent and later less frequent. Further, new strategies are introduced and old strategies stop being used. This overlapping waves model seems more in accord with what is known about cognitive development than do depictions that show children suddenly moving from one approach to another.

My colleagues and I have pursued this evolutionary model within a variety of areas: arithmetic, time telling, reading, spelling, problem solving, and memory tasks, among them (e.g., Crowley & Siegler, 1993; Siegler, 1996; Siegler & Strager, 1984). In each of these areas, we have found that competition leads to adaptive consequences, and that basic strategy choice and discovery mechanisms produce the adaptation. The findings can be illustrated in the context of young children’s learning of simple addition.

FIGURE 3.7 Siegler’s overlapping waves model of cognitive development.
First consider the competing entities. Even 5-year-olds use a variety of strategies to solve basic addition problems such as 3 + 5. Sometimes they count from one; this typically involves putting up fingers on one hand to represent the first addend, putting up fingers on the other hand to represent the second addend, and then counting the raised fingers on both hands. Other times, they put up fingers but recognize the number of fingers that are up without counting. Yet other times, they retrieve an answer from memory. Some children also know another strategy, the count-on strategy. Children using this strategy choose the larger of the two addends and count on from that point the number of times indicated by the smaller addend. For example, for 3 + 9, children might think to themselves, "9, 10, 11, 12."

It is not the case that some 5-year-olds use one of these strategies and some use another. Rather, almost all use several different strategies. Similarly, on subtraction, multiplication, spelling, time telling, and memory tasks, the majority of children have been found to use at least three strategies. Even on individual problems, the outcomes of the competition vary, so that the same child will choose one strategy one day and a different one the next (Siegler, 1987a).

Children's choices among these strategies are adaptive in several different ways. One sense in which their choices are adaptive is that they use retrieval, the fastest strategy, predominantly on simple problems where it can yield accurate performance, and use more time-consuming and effortful strategies on more-difficult problems, where such strategies are necessary for accurate performance (Siegler, 1986).

Children also choose adaptively among strategies other than retrieval. In particular, they tend to use each strategy most often on problems where it works especially well compared to alternative approaches. In evolutionary terms, strategies find their niches. For example, the counting-on strategy is used most often on problems such as 2 + 9, where the smaller addend is quite small and the difference between addends is large. On such problems, counting-on is both easy to do and effective relative to alternative procedures such as counting-from-one (Siegler, 1987b).

Changes over time in strategy use also are adaptive. For example, in simple addition, children increasingly use the most efficient strategies, such as retrieval and counting on, and decrease their use of less-efficient strategies, such as guessing and counting from one. They also acquire new strategies, such as decomposition (e.g., solving 3 + 9 by thinking "3 + 10 = 13, 9 is 1 less than 10, so 3 + 9 = 12").

What type of selection mechanisms could produce such adaptive strategy choices? The model that I formulated divides the information-processing system into representations and processes. The representations include factual information and data; the processes operate on the representations to produce behavior. For example, in the context of arithmetic, the representation includes associations between problems and various possible answers to the problems. The
processes are strategies such as counting from one, counting on, and retrieval that solve problems by operating on the data in the representation.

Figure 3.8 illustrates how this type of organization could yield effective choices among strategies at any one time and adaptive changes in strategy use over time. Within the model, the use of strategies to solve problems generates answers to the problems and also generates information about the speed and accuracy with which the problem was solved. This information feeds back to provide increasingly detailed knowledge about both the strategies and the problems. Subsequent choices among strategies are made on the basis of their past effectiveness in solving problems in general, in solving particular kinds of problems, and in solving specific problems. The more effective a strategy has been in solving problems in the past, the more often it will be chosen in the future. Further, the choices among strategies become increasingly refined as children learn that a strategy that in general is the most effective is not necessarily the most effective for a particular type of problem.

This view of development has provided the basis for computer simulations of the development of arithmetic (Siegel & Strayer, 1984; Siegel & Shipley, 1989). To illustrate how the general theoretical assumptions are realized within a specific simulation, I will describe Siegel and Shipley's (1995) model of the development of single-digit addition. This simulation modeled how children learn to choose among three approaches: counting from one, counting on from the larger addend, and retrieval. Its working can be illustrated by considering its strategy choices on 9 + 1. The simulation gradually learns that it is easier to solve this problem by counting on from the larger addend than by counting from one. It requires far fewer counts to say "9, 10" than "1, 2, 3, 4, 5, 6, 7, 8, 9, 10." This lesser amount of counting results in fewer errors and shorter solution times, which, in turn, leads to more frequent future choices of the counting-on strategy. The simulation uses this experience and similar experience with other

![Figure 3.8](image-url)
problems to draw the generalization that counting on works better than counting from one on related problems, such as \( 9 + 2 \) and \( 8 + 1 \) and uses the knowledge to generalize appropriately to unfamiliar problems.

As children increasingly choose strategies that correctly solve problems, they also increasingly associate the correct answers with the problems. For example, \( 9 + 1 \) becomes strongly associated with \( 10 \). This association allows them to retrieve \( 10 \) as the answer to the problem. Retrieving an answer is even faster than counting "9, 10" and is just as accurate. Thus, the very success of the counting-on strategy in producing correct answers leads to its own obsolescence, because it makes accurate retrieval possible.

The evolutionary perspective raises the further issue of the source of strategic variation. In particular how are new strategies acquired? Sometimes children are taught a new strategy or imitate another person who is using it. However, the most interesting case is strategy discovery, in which children invent a strategy for themselves.

How do children discover new strategies? To find out, Siegler and Jenkins (1989) examined 4- and 5-year-olds’ discovery of the counting-on strategy. Recall that this strategy involves solving problems such as \( 2 + 9 \) by thinking, "9, 10, 11." At the beginning of the study, children in the Siegler and Jenkins experiment knew how to add by counting from one, but did not yet know how to do so by counting on from the larger addend. The children practiced solving addition problems three times per week for 12 weeks. Because even young children can accurately report immediately after an addition problem how they solved the problem (Siegler, 1987), it was possible to identify the exact trial on which each child first used the new strategy. This allowed us to examine what led up to the discovery, what the experience of discovering a new strategy was like for a child, and how children generalized the new strategy to other problems after making the discovery.

Almost all of the children discovered the new strategy during the course of the experiment. The time that they took to make the discovery varied widely; the first discovery came in the second session, whereas the last one did not come until the thirtieth session. The quality of the discoveries also varied widely. Some discoveries showed a great deal of insight, as exemplified by "Lauren’s" protocol:

E: How much is 6 + 3?
L: (Long pause) Nine.
E: OK, how did you know that?
L: I think I said ... I think I said ... oops, um ... I think he said ... 8 was 1 and ... um ... I mean 7 was 1, 8 was 2, 9 was 3.
E: How did you know to do that? Why didn’t you count 1, 2, 3, 4, 5, 6, 7, 8, 9?
L: Cause then you have to count all those numbers.
E: OK, well how did you know you didn't have to count all of those numbers?

L: Why didn't? . . . well, I don't have to if I don't want to. (Siegler & Jenkins, 1989, p. 66)

Other children did not show nearly as much understanding. In fact, some claimed not to have counted at all, despite videotapes of their performance providing audible and visible evidence that they had used the new strategy.

What led up to the discoveries? Our expectation had been that difficult problems, or situations in which children failed to solve previous problems, would elicit them. This proved not to be the case, however. The problems on which discoveries were made, and the accuracy of performance just prior to discovery, did not differ from performance in the rest of the experiment. The only distinguishing characteristic of performance immediately before the discovery was solution times that greatly exceeded the usual amount. For example, Lauren, the child quoted above, took 67 seconds to generate the answer on the trial just before her discovery and 35 seconds on the trial where she first used the new strategy. Both trials were much longer than her average time of 11 seconds. These long times were accompanied by numerous false starts, pauses, odd statements (as in Lauren's comment referring to her own counting "I think he said"), and other indicators of cognitive ferment.

Another striking characteristic of the children's performance was how slowly they generalized the new strategy to other problems. For example, one girl used the new strategy on only 7 of the first 64 problems after her discovery; another girl did so on only 2 of the first 49. Temporarily, the children continued to rely heavily on the familiar counting-from-one approach, even though they knew the potentially more effective count-on strategy.

The amount of generalization increased dramatically, however, when challenge problems were presented to the children in the eighth week of the study. These were problems such as 24 + 2 that were easy to solve using counting on but almost impossible to solve using counting from one or retrieval. For children who had discovered the strategy of counting on from the larger addend, encountering these problems seemed to increase awareness both of the new strategy as constituting a different approach and of the goals that the new strategy could meet. In a sense, they rediscovered the strategy in a way that increased their understanding of how it served the goals of addition. After encountering the challenge problems, they generalized the counting-on strategy much more widely than they had before, on small as well as large number problems. On the other hand, children who had not used the counting-on strategy at all did not benefit from being presented the challenge problems; they simply became confused.

What are the main limitations of this theory? One problem is that the mechanisms that produce variation have not been spelled out in the same detail as the mechanisms that produce selection. Whereas the computer simulations of strategy choice indicate in detail how choices are made among existing strategies.
and how choices among the strategies change over time, no similarly detailed model exists for how new strategies are discovered. Also, the theory seems most applicable to domains in which children use clearly-defined strategies; its applicability to areas in which strategies are less well defined remains to be demonstrated. Still, it would be disingenuous of me to appear pessimistic about it. The basic observation that cognitive development resembles biological evolution is beginning to emerge in many areas: perceptual development (Johnson & Karmiloff-Smith, 1992), language development (MacWhinney & Chang, 1995), motor development (Thelen, 1995), and analogical reasoning (Gentner, 1989) among them. If the approach proves half as useful in understanding cognitive development as it has in understanding biological evolution, the effort to apply the idea will be well worthwhile.

**Developmental mechanisms work together.** In this chapter, contributions to cognitive development of the four mechanisms described earlier—automatization, encoding, generalization, and strategy construction—have been discussed in the contexts of different theories. This reflects the fact that different theories emphasize different mechanisms. However, all of the theories recognize that all of these mechanisms (and others) work together to produce cognitive change.

Consider an example of how all four mechanisms might contribute to a single development—a girl learning to attach "ed" to verbs to indicate that the action occurred in the past. Early in the process of language development, all of her mental resources would be needed just to perceive clearly words and phrases she heard. (Think of hearing a conversation in a foreign language that you don’t know or are just starting to learn.) With greater experience listening to people talk, her processing of the words and phrases would become automatized, freeing up cognitive capacity for other types of processing. This extra capacity would allow her to notice that similar meanings were often expressed by words that ended with an ed sound. This realization, in turn, would lead her to encode the ed sound as a separate unit, to find out just what it meant. She then could note the regular connection between the ed sound and the action having occurred in the past. Finally, she could construct a new strategy based on the generalization: Whenever you want to indicate that an action occurred in the past, attach an ed to the end of the word describing the action. By working together in this way, automatization, encoding, generalization, and strategy construction may account for many improvements in children’s thinking.

**SUMMARY**

Information-processing theories of development have several distinguishing characteristics. Their basic assumption is that thinking is information processing. They emphasize precise analysis of change mechanisms. They focus on the
strategies that children devise to surmount the challenges posed by the environment and by their own limited processing capacity and knowledge.

Within information-processing approaches, cognition is viewed as reflecting both structure and process. Structure refers to relatively fixed aspects of the information-processing system, process to relatively variable and changeable ones. Among the most critical structures are sensory, working, and long-term memory. Sensory memory is devoted to holding a relatively large amount of unanalyzed information for about a second after the information is encountered. Working memory involves the information in the current situation and in long-term memory that is receiving attention at any given time. Without continuing attention, information is lost from working memory within 15 to 30 seconds. Long-term memory involves our enduring knowledge of procedures, facts, and specific events. It appears to be of unlimited capacity, and information remains in it indefinitely.

In contrast to this relatively small number of structures, each of which influences thinking in almost all situations, a much larger group of processes contributes in more delimited situations. These processes vary greatly with the particular circumstances, thus giving human cognition much of its flexibility. The same situation also elicits different processes in different people, depending on their past experience and abilities. Rules, concepts, and strategies are among the types of processes that people most often use.

Several information-processing theories of development have been formulated to make understandable how creatures as helpless and ignorant as infants eventually attain the power and flexibility of the adult information-processing system. Neo-Piagetian theories are aimed at uniting Piagetian and information-processing theories. Case’s approach is a particularly influential example. It posits a series of stages much like Piaget’s and a set of central conceptual structures that organize thinking in domains such as number, space, and narratives. It also suggests that limited working-memory capacity is a major obstacle to cognitive growth. By automatizing their processing, through biological maturation and through acquisition of more advanced central conceptual structures, children become able to perform increasingly difficult cognitive feats.

Psychometric theories are intended to reveal the processes underlying the individual differences that appear on intelligence tests. Sternberg’s triarchic theory of intelligence illustrates how information-processing ideas can be used to pursue this goal. Sternberg divides intelligence into three types of components: metacomponents, performance components, and knowledge-acquisition components. Metacomponents function as a strategy-construction mechanism, arranging the other two types of components into goal-oriented procedures. Knowledge-acquisition components are used to obtain new information when no solution to a problem is immediately possible. Performance components do the work of solving the problem. The theory has been applied to diverse cognitive skills and to many populations, including gifted and retarded children.
Production-system theories are intended to explain how changes in problem solving occur. Klahr’s theory explains particularly clearly how self-modifying production systems can advance understanding of development. It focuses on the developing system’s capacity for generalization. In this analysis, generalization includes three components: the time line, regularity detection, and redundancy elimination. The time line is a record of all the situations the system has encountered, its responses to the situations, and the outcomes. Regularity detection operates on the data in the time line to detect repeated patterns. Redundancy elimination looks for parts of procedures that could be eliminated without changing the outcome of processing. Together, these mechanisms allow children to generalize their knowledge to new situations.

Connectionist theories are a class of computer simulation models based on an analogy to the workings of the brain. In them, numerous simple processing units, analogous to neurons, are connected to each other with varying strengths. When presented input, the processing units receive activation from each other, with the processing activity leading to a response. The response is compared to the correct answer, and the strengths of connections are adjusted in ways that would have led to more accurate responding. MacWhinney et al. demonstrated how such a model could learn the German language’s complex system for determining whether given nouns are masculine, feminine, or neuter and which article should be attached to a given noun. The system’s learning resembled that of children both in which nouns were hardest to learn and in the types of errors that it made.

Evolutionary theories are based on an analogy between biological and cognitive evolution. As emphasized in my own approach, the critical contributors to change in both cases are sources of variation and sources of selection. In children’s thinking, strategy discovery provides one source of variation; strategy choice procedures provide a means of selection. The two types of processes work together to change not only how often children use different strategies, but also when they use each approach. The theory has stimulated observations of how children construct new strategies and of how use of existing strategies changes over time.

RECOMMENDED READINGS


