Variation, Selection, and Cognitive Change

Robert S. Siegler
Christopher Shipley
Carnegie Mellon University

Cognitive development is usually described in terms of a sequence of 1:1 relations between ages and ways of thinking. This is evident in both classical domain-general stage theories and in recent domain-specific approaches. First, consider the stage theories. Within Piaget's theory, the reasoning of young children is said to be preoperational; that of somewhat older ones, concrete operational; that of yet older ones, formal operational. Within Vygotsky's theory, young children are said to form thematic concepts; somewhat older ones, chain concepts; yet older ones, true concepts. Within Bruner's theory, infants are said to form sensorimotor representations; somewhat older children, iconic representations; yet older ones, symbolic representations.

Current cognitive development has moved away from many aspects of stage theories, particularly their domain generality and their overly conservative estimates of infants' and young children's cognitive capacities. One feature that has been retained, however, is the emphasis on 1:1 relations between ages and ways of thinking. In descriptions of the development of competence in forming past tense verbs, children are said first to use correctly a small range of regular and irregular past tense forms, then to overgeneralize the standard past tense rule to irregular as well as regular verbs, then to form correctly both regular and irregular verbs (Brown, 1973; Rumelhart & McClelland, 1986). In descriptions of the development of addition skill, 5-year-olds are said to count from 1; 7-year-olds to count from the larger addend; 10-year-olds to retrieve answers from memory (Ashcraft, 1982; Neches, 1987). In descriptions of the development of serial
recall strategies, 5-year-olds are said not to rehearse; 8-year-olds to rehearse in a simple way; 11-year-olds to rehearse in a more elaborated way (Flavell, Beach, & Chinsky, 1966; Schneider & Pressley, 1989).

Although these 1:1 equations between ages and ways of thinking are omnipresent in the literature, few would defend them as literally meaning that young children of a given age or developmental level always use one approach, older ones always use another approach, and so on. Instead, their widespread and enduring use seems due to their having several pragmatic advantages: They are interesting, sometimes dramatic, easy to describe, easy to remember, and straightforward to discuss in textbooks and lectures.

The 1:1 equations also entail two serious problems, though, one obvious, the other less so. The obvious problem is that the 1:1 equations are inaccurate. For example, detailed studies of each of the previously described domains show that individual children, by definition of a single age, generate a variety of past tense verb forms, arithmetic strategies, and serial recall strategies (Maratsos, 1983; McGilly & Siegler, 1990; Siegler & Robinson, 1982). Such inaccuracy is a serious problem. Still, if it were the only one, many might still judge the trade-off worthwhile.

A less obvious but equally pernicious consequence of this oversimplification is that it impedes understanding of change. The 1:1 equations between age and way of thinking make it extremely difficult to go beyond a superficial understanding of change. How can preoperational reasoning turn into concrete operational reasoning, chain concepts into true concepts, iconic concepts into symbolic concepts? How can a child who always overregularized past tense verbs become one who correctly uses both regular and irregular forms? Portraying children's thinking as monolithic at each point in the developmental sequence has the effect of segregating change from the ebb and flow of everyday cognitive activity. It makes change a rare, almost exotic, event that demands an exceptional explanation. Thus, we often speak of brief "transition periods" that separate long-lived stages, theories, strategies, or rules. Yet, if children of a given age have for several years been in a particular stage, had a particular theory, or used a particular strategy or rule, why should they suddenly change? From this perspective, it is no accident that both traditional and current depictions of development have been so consistently faulted for being vague, unclear, mysterious, or silent concerning how transitions occur. The problem is that the 1:1 equations that the accounts of change seek to explain are fundamentally flawed, root and branch.

The central theme of this chapter is that recognizing the roles of variation and selection in cognitive development, and transcending the omnipresent 1:1 equation between age and thought, will lead to a better understanding of how change occurs. We illustrate this perspective in the context of
children's strategy choices, particularly their choices among alternative strategies for adding numbers. We first note key empirical phenomena that any model of strategy choice must explain. Then we describe three generations of models in this area and their successes and failures in accounting for the empirical phenomena. Finally, we consider their contributions to understanding how change occurs and the issues they raise for future investigation.

BASIC PHENOMENA

Variability

Although innumerable studies have depicted development in terms of a 1:1 correspondence between children's age and the strategy they use, recent trial-by-trial analyses indicate that children of a single age often use a variety of strategies. One task in which we observed such strategy discovery is elementary school children's single-digit addition. Examination of both videotaped records of ongoing behavior and immediately retrospective self-reports reveals five relatively common strategies (each used on 3% to 36% of trials among the kindergartners, first and second graders in Siegler, 1987b). Sometimes children use the sum strategy, in which they count from one; to solve $3 + 6$, a child using this strategy might put up 3 fingers, then 6 more, and then count from 1 to 9. Other times, children use the min strategy, which involves counting from the larger addend the number of times indicated by the smaller addend. Here, a child would solve $3 + 6$ by counting "6, 7, 8, 9" or "7, 8, 9." On other occasions, children use decomposition—translating the problem into an easier form and then making the necessary adjustment. A child solving $3 + 6$ via decomposition might think, "$3 + 7 = 10$, 6 is 1 less than 7, so $3 + 6 = 9."$ Still other times, children use retrieval or guessing to generate an answer. These strategies are described in Table 2.1.

These diverse strategies are not artifacts of one child consistently using one of the strategies and another child consistently using a different one. The majority of kindergartners, first and second graders use at least three of the five strategies; a substantial minority use more (Siegler, 1987b). This multiple strategy use is apparent within classes of similar problems, and even on the same problem presented to the same child on consecutive days. In two studies, one on addition (Siegler & Shrager, 1984), and one on time telling (Siegler & McGilly, 1989), fully one-third of the children used different strategies on the identical problem on two successive days. Only a small part of this day-to-day variability could be explained by learning, because the progression of strategies was not consistently from less to more
TABLE 2.1
Early Elementary School Students' Main Addition Strategies

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Typical Use of Strategy to Solve 3 + 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum</td>
<td>Put up 3 fingers, put up 5 fingers, count fingers by saying &quot;1, 2, 3, 4, 5, 6, 7, 8.&quot;</td>
</tr>
<tr>
<td>Min</td>
<td>Say &quot;5, 6, 7, 8&quot; or &quot;6, 7, 8,&quot; perhaps simultaneously putting up one finger on each count beyond 5.</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Say an answer and explain it by saying, &quot;I just knew it.&quot;</td>
</tr>
<tr>
<td>Guessing</td>
<td>Say an answer and explain it by saying, &quot;I guessed.&quot;</td>
</tr>
<tr>
<td>Decomposition</td>
<td>Say &quot;3 + 5 is like 4 + 4, so it's 8.&quot;</td>
</tr>
</tbody>
</table>

advanced. For example, in the study of addition, almost as many children retrieved the answer on the first day, and used the sum strategy on the second, as did the reverse (45% vs. 55%).

This strategy diversity is not limited to any particular domain or age group. Consider findings from our own studies. When multiplying, 8- to 10-year-olds sometimes repeatedly add one of the multiplicands, sometimes write the problem and then recognize the answer, sometimes write and then count groups of hatch marks that represent the problem, and sometimes retrieve the answer from memory (Siegler, 1988b). To tell time, 7- to 9-year-olds sometimes count forward from the hour by ones and/or fives, sometimes count backwards from the hour by ones and/or fives, sometimes count from reference points such as the half hour, and sometimes retrieve the time that corresponds to the clock hands' configuration (Siegler & McGilly, 1989). To spell words, 7- and 8-year-olds sometimes sound out words, sometimes look them up in dictionaries, sometimes write out alternative forms and try to recognize which is correct, and sometimes recall the spelling from memory (Siegler, 1986). To serially recall lists of unrelated stimuli, 5- to 8-year-olds sometimes repeatedly recite the names of items within the list during the delay period, sometimes recite the names once and stop, and sometimes just wait (McGilly & Siegler, 1989, 1990).

Similar strategy diversity has been observed by other investigators of arithmetic (Cooney, Swanson, & Ladd, 1988), causal reasoning (Shultz, Fisher, Pratt, & Rulf, 1986), number conservation (Church & Goldin-Meadow, 1986), spatial reasoning (Ohlsson, 1984), referential communication (Kahan & Richards, 1986), language development (Kuczaj, 1977), and motor activity (Goldfield, in press). These diverse strategies have been observed among adults as well as children, and in Japan and China as well as in North America and Europe (Geary, Fan, & Bow-Thomas, 1992; Kuhara-Kojima & Hatano, 1989). These studies call into question the typical description that at Age N, children are in Stage X, have Rule X, have Theory X, use Strategy X; at Age N + 2 they are in Stage Y, have Rule Y, have Theory Y, use Strategy Y; and so on. Instead, over a wide range of ages and tasks, children know and use multiple approaches.
Adaptive Strategy Choices

Children's strategy choices are adaptive in several ways. One involves their choice of whether to state a retrieved answer or to use a backup strategy (any approach other than retrieval, such as the sum and min strategies in addition). As shown in Fig. 2.1, the more difficult the problem, the more often children use backup strategies to solve it. This pattern of strategy choice is adaptive, because it enables children to use the faster retrieval approach on problems where that yields correct answers and to use the slower backup strategies on problems where they are necessary to produce accurate performance. Consistent with this analysis, forcing children to retrieve on all trials by imposing a short time limit (4 sec) produces a sharp fall-off in accuracy, with the fall-off largest on precisely the problems on which children most often use backup strategies when allowed to choose freely (Siegler & Robinson, 1982).

This pattern of choices between retrieval and backup strategies is extremely general. It holds true with problem difficulty defined either by percentage of errors or by length of solution times; with preschoolers, elementary school-aged children, college students, and senior citizens; with high achieving and low achieving students; with suburban White and inner-city African-American children; and with addition, subtraction, multiplication, time telling, spelling, and word identification (Geary & Burgham-Dubree, 1989; Geary & Wiley, 1991; Kerkman & Siegler, 1993; Maloney & Siegler, 1993; Siegler, 1986, 1988a, 1988b; Siegler & Shragon, 1984).

Children also choose adaptively among alternative backup strategies. For example, when choosing between the min and sum strategies, children most often select the min strategy on problems where differences between addends are large (Siegler, 1987b). Thus, when solving single-digit addition problems, they are most likely to use the min strategy on problems such as 9 + 2. Problems with large differences between addends are the ones on which the min strategy produces the greatest savings in amount of counting, relative to the main alternative approach, the sum strategy; thus, it makes sense to use it most often on them.\(^1\)

Change

Three main changes in strategy use occur with age and experience: changes in relative frequency of use of existing strategies, changes in the effective-

\(^1\)It might seem that using the min strategy would always be more advantageous than using the sum strategy, because it always involves less counting. However, at the time when they learn the min strategy, children are much more practiced at counting from one than from other starting points, and do so more accurately and efficiently. Thus, the advantage of the min over the sum strategy is not as great as the savings in number of counts would suggest, and some problems at first are solved more quickly via the sum strategy (Siegler & Jenkins, 1989).
FIG. 2.1. Relations between percent errors on each problem and percent use of backup strategies on that problem on four tasks.

ness with which each strategy is executed, and changes involving acquisition of new strategies.

Table 2.2 illustrates the changes in strategy use that occur in kindergartners', first, and second graders' single-digit addition. Frequency of retrieval, the min strategy, and decomposition increase; frequency of the sum strategy and guessing decrease.

Changes in the skill with which each strategy is executed are also evident.
2. VARIATION, SELECTION, AND CHANGE

TABLE 2.2
Percent Use of Each Strategy by Children of Each Age

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>Strategy</th>
<th>Retrieval</th>
<th>Min</th>
<th>Decomp</th>
<th>Count all</th>
<th>Guess or no response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten</td>
<td></td>
<td>16</td>
<td>30</td>
<td>2</td>
<td>22</td>
<td>30</td>
</tr>
<tr>
<td>Grade 1</td>
<td></td>
<td>44</td>
<td>38</td>
<td>9</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Grade 2</td>
<td></td>
<td>45</td>
<td>40</td>
<td>11</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>


Older children are faster and more accurate than younger ones in their execution of each addition strategy. For example, in Siegler (1987b), mean time to execute the min strategy decreased from 6 sec among kindergartners to 4 sec among second graders, and percent errors using the strategy decreased from 29% to 7%. Thus, the general improvements in speed and accuracy that characterize acquisition of arithmetic include both changes in frequency of use of different strategies and in efficiency of execution within each strategy.

A third type of change involves acquisition of new strategies. Even after children know strategies that consistently yield correct performance on a class of problems, they continue to invent new strategies for solving them. For example, children who competently solve simple addition problems by using the sum strategy or retrieval go on to discover the min strategy (Siegler & Jenkins, 1989). In class inclusion, children who competently solve problems by comparing the number of objects in the larger subordinate class with the number in the superordinate class later realize that the superordinate class must be larger, whatever the particular numbers involved (Markman, 1979). In using causal verbs, deeper linguistic analysis leads children who for years have made such grammatically correct statements as "she killed him" and "he dropped it," to start introducing ungrammatical forms such as "she died him" and "he failed it" (Bowerman, 1982). Together with imitation and direct instruction from other people, such strategy discovery is among the ultimate sources of variability in children's strategy use.

Generalization

Children generalize their strategies beyond the particular problems on which they acquired them. Observation of children in the weeks after they first discover a new strategy indicates that this generalization process often
takes a long time, at least when children possess other strategies that work well (Kuhn, Schauble, & Garcia-Mila, 1992; Schauble, 1990; Siegler & Jenkins, 1989). However, generalization of new strategies can be considerably hastened if children encounter conditions that highlight the new strategy's advantages over previous approaches.

Both the typical slow generalization and a condition that produced much faster generalization were observed in a study of 4- and 5-year-olds' discovery of the min strategy (Siegler & Jenkins, 1989). At the beginning of the experiment, the preschoolers knew how to add via the sum strategy and retrieval, but did not yet know the min strategy. They were then presented with 7 weeks of practice on small number problems, where neither addend exceeded 5. During this period, most of the children discovered the min strategy, in the sense of using it at least once. These early uses were often accompanied by insightful explanations of why the strategy was useful (e.g., "If you count from 4, you don't need to count all those numbers").

Despite these insights, the children showed little generalization of the newly discovered strategy, instead continuing to rely mainly on the sum strategy when they could not retrieve the answer. Therefore, in Week 8, they were presented with challenge problems, such as $3 + 22$, that included a small and a very large addend. The logic was that such problems would serve as both a carrot, encouraging use of the min strategy, and as a stick, discouraging use of the sum strategy or retrieval.

Children who had not yet discovered the min strategy did not benefit in any way from encountering such challenge problems; most could not cope with them. However, those who had previously discovered the min strategy (in the sense of using it at least once) generalized it much more widely than they had before. The amount of generalization continued to increase in the three remaining weeks of the experiment, on small addend problems as well as on the challenge problems. By the end of the experiment, the min strategy was the dominant backup strategy, being used on more than 90% of trials where retrieval was not. Both the slow generalization that occurs in the absence of experiences that highlight the advantages of new strategies, and the rapid generalization that can take place when such experiences occur, are important phenomena for models of strategy choice to explain.

Individual Differences

Recognizing the variability of strategy use within individuals raises the issue of whether strategy use also varies in interesting ways across individuals. Although research on broad cognitive styles has not identified many strong consistencies in strategy use (Kogan, 1983; Sternberg, 1985), studies of more narrowly defined strategy choices have yielded more encouraging results. For example, examination of the choice between stating a retrieved
answer and using a backup strategy revealed consistent individual patterns in first graders' addition, subtraction, and word identification (Kerkman & Siegler, 1993; Siegler, 1988a). The research delineates three characteristic patterns: the good student, not-so-good student, and perfectionist patterns. Good students are children who usually rely on retrieval and answer quickly and accurately. Not-so-good students sometimes use retrieval but usually answer slowly and inaccurately. Perfectionists are fairly fast and very accurate, but use retrieval even less often than the not-so-good students; instead, they rely heavily on backup strategies.

These patterns of individual differences in addition, subtraction, and word identification have been found in four different experiments, some involving high income, predominantly White, suburban children, and others involving low income, predominantly African American inner-city children. They also have proved predictive of standardized test scores and of future classroom placements; not-so-good students score significantly lower than the other two groups, are more likely to be classified as learning disabled, and are more likely to need to repeat a grade (Kerkman & Siegler, 1993; Siegler, 1988a; Siegler & Campbell, 1990). Thus, the differences between not-so-good students and the other two groups are of the type that are detected by standard psychometric tests. However, the differences between the strategy choices of good students and perfectionists are not apparent on these tests; they are more akin to cognitive style differences.

Summary

The present perspective brings to center stage a different set of phenomena than those emphasized in traditional developmental accounts. Rather than focusing on the problem solving approach children use at each age, our perspective focuses on the set of approaches children use. Highlighting this variability brings into the spotlight how children choose among the alternative approaches and what adaptive purposes those choices serve. Our perspective also calls attention to several different types of changes: changes in the frequency of use of existing strategies, in the effectiveness with which these strategies are used, and in the acquisition of new strategies. Beyond this, our perspective spotlights how strategies are generalized beyond their initial contexts, and how individuals vary in their strategy choices. We now describe and evaluate three generations of models that have attempted to account for these phenomena.

**GENERATION 1: METACOGNITIVE MODELS OF STRATEGY CHOICE**

Children's strategies first became a major topic of research in the 1960s (e.g., Flavell, Beach, & Chinsky, 1966; Keeney, Cannizzo, & Flavell, 1967).
The early research documented large changes in memory strategies between ages 5 and 8. Five-year-olds were said to rarely use strategies such as rehearsal and organization; 8-year-olds were said to consistently use them. Especially interesting were efforts to teach such strategies to 5- and 6-year-olds who did not spontaneously use them. Such children often learned the strategies, and their performance improved when they used them. Despite these benefits, the children usually did not continue to use the strategies later, even in similar situations.

This puzzle was an important impetus for the first generation of models of strategy choice. These models were labeled metacognitive because they focused on how knowledge about cognition could be used to control cognitive activities. Their fundamental assumption was that young children's failure to use new strategies reflected their limited understanding of their own cognitive capacities and of why the new strategies were needed.

The metacognitive models originally proposed, the type being discussed in the next section, focus on explicit, rationally-derived, conscious metacognitive knowledge. The term metacognitive also at times has been used to refer to processes that are implicit, not derived from rational consideration, and unconscious. Greeno, Riley, and Gelman's (1984) planning networks, Van Lehn's (1982) repair models, and Halford's (1993) mental models, like our own models of strategy choice, fit into this category. In the present context, however, the term metacognitive is used only in its original sense of explicit, rationally-derived, conscious knowledge that influences the workings of the cognitive system. This allows a clearer contrast between the two approaches to strategy choice.

Two Metacognitive Models

Metacognitive approaches assume that strategy choices are made through the cognitive system's explicit knowledge of its own workings. This knowledge is often said to be used by an executive processor, that decides what the cognitive system should do (Case, 1978; Kluwe, 1982; Sternberg, 1985). Schneider and Pressley (1989) described the executive processor's role as follows:

This executive is aware of the system's capacity limits and strategies. The executive can analyze new problems and select appropriate strategies and attempt solutions. Very importantly, the executive monitors the success or failure of ongoing performance, deciding which strategies to continue and which to replace with potentially more effective and appropriate routines. In addition, the efficient executive knows when one knows and when one does not know, an important requirement for competent learning. (p. 91)
Similarly, Kuhn (1984) described the way in which metacognitive knowledge influences strategy selection as follows: "In order to select a strategy as appropriate for solving a particular problem, the individual must understand the strategy, understand the problem, and understand how the problem and strategy intersect or map onto one another" (p. 165).

Models of the executive processor that are based on such conceptions of how metacognitive knowledge exercises its effects have been high level, rather abstract characterizations of types of relevant knowledge used to govern cognition. A less and a more elaborate model of this type are shown in Fig. 2.2.

Such metacognitive models are useful for conveying hypotheses about relations among different types of knowledge and for pointing to one way in which intelligent strategy choices can be generated. However, they also have a number of weaknesses, both theoretical and empirical (Brown & Reeve, 1986; Cavanaugh & Perlmutter, 1982; Siegler, 1988b). As statements of theory, they generally have been vague regarding the mechanisms that produce the phenomena of interest. Do people make explicit judgments about their intellectual capacities, available strategies, and task demands every time they face a task they could perform in multiple ways? If not, how do they decide when to do so? Do they consider every strategy they could use on the task, or only some of them? If only a some, how do they decide which ones? How do people know what their cognitive capacity will be on a novel task or what strategies they could apply to it? The apparent simplicity of metacognitive models masks a world of complexity.

Empirical evidence has also raised questions about the fundamental assumption that underlies the models. Relations between explicit, verbalizable metacognitive knowledge and cognitive activity have proven weaker than originally expected (Cavanaugh & Perlmutter, 1982; Schneider, 1985; Schneider & Pressley, 1989), casting doubt on whether such metacognitive knowledge plays a central role in children's strategy choices.

On the other hand, the questions addressed by metacognitive research are important, and research stimulated by the approach has yielded intriguing data. We now review findings from this research that are relevant to the five phenomena noted in the previous section.

Variability

Metacognitive research has documented that even quite young children have conscious, stabile knowledge about diverse strategies. For example, Kreutzer, Leonard, and Flavell (1975) asked 5- to 10-year-olds what they could do to remember to bring their skates to school the next day. At all ages, children often described multiple strategies. Studies of other tasks
FIG. 2.2. Two metacognitive models of the executive processor: Flavell's (1981) model (top) and Kluwe's (1982) model (bottom).
(e.g., ways of remembering a forgotten idea) have obtained similar results at young ages and have demonstrated that the number of strategies described increases at least into early adolescence (Yussen & Levy, 1977).

Adaptive Strategy Choices

The metacognitive perspective suggests a way of thinking about how children choose among alternative strategies. The models are often not explicit, but the implicit causal pathway is:

Metacognitive knowledge → Strategy choice → Performance

For example, knowledge about alternative memory strategies, task demands, and personal capacities would lead the executive to choose a particular memory strategy, that in turn, would influence memory performance.

The type of research suggested by this approach is exemplified in an experiment reported by Justice (1985). Children were presented with videotapes of a 10-year-old executing different strategies that might benefit recall, and were then asked to judge the likely effectiveness of the strategies. Older children's judgments of the strategies' relative value were more accurate. The conclusion was that older children's superior choice of strategies and superior memory performance largely stemmed from their superior knowledge about the strategies.

However, even when children know abstractly that one strategy is superior to another, they often do not choose that strategy. Flavell and Wellman (1977) and Wellman (1983) suggested a number of explanations for this discrepancy: for example, lack of motivation, lack of time, overreliance on sheer effort, and beliefs that the strategy is unnecessary. Supporting this view, Fabricius and Hagan (1984) found that among a sample of children taught a new strategy that consistently improved their performance, only a minority continued to use it when no longer instructed to do so. The children who continued to use it were predominantly those who attributed their improved success to adopting the strategy. Other children, who attributed their improvement to trying harder, luck, or other factors, generally did not continue to use the new approach.

Even when significant relations between metacognitive knowledge and strategy choices are present, they often do not seem sufficient to account for the strong relations between problem difficulty and strategy choices that have been observed (Siegler, 1986). This can be seen in the specific area of focus in this chapter, children's arithmetic. Recall that the frequency with which children choose to use backup strategies on a given arithmetic problem is highly correlated with the problem's difficulty. This relation
might have been mediated by metacognitive judgments of problem difficulty. Actual problem difficulty could give rise to metacognitive judgments of problem difficulty, which in turn could direct strategy choices. A difficult problem could lead a child to think, “This is a difficult problem; I’d better use a strategy such as X that can solve problems like that.” However, Siegler and Robinson (1982) found that children’s judgments of the difficulty of a set of arithmetic problems correlated only $r = .51$ with their strategy choices (the percentage of trials on which they used backup strategies on the problem). Further, the metacognitive judgments of problem difficulty correlated only $r = .47$ with actual problem difficulty (measured as percent errors the problem elicited). These correlations were significant, but not nearly sufficient to account for the very strong correlation ($r = .92$) between problem difficulty and strategy choices (25% vs. 85% variance accounted for). Young children’s moderate amount of conscious, explicit, metacognitive knowledge may contribute to their extremely adaptive strategy choices in arithmetic and other domains, but cannot alone account for them.

Change

Metacognitive knowledge increases greatly during childhood and adolescence. Older children know more strategies, are often better at choosing among them, and are better at learning new approaches as well (Schneider & Pressley, 1989). They are more realistic in assessing their own memory capacities, more accurate in assessing the relative importance of different parts of a task, and more knowledgeable about interactions among factors that influence performance. This increasing metacognitive knowledge provides a database on which children increasingly can rely to choose strategies. Whether reliance on such knowledge to choose among strategies increases with age, however, remains unknown.

Generalization

When children learn a strategy, they often do not generalize it to new situations. One reason may be that they often realize fewer benefits in increased accuracy and incur greater costs in cognitive effort from using new strategies than they will when the strategies become better practiced (Gutentag, 1984; Miller, 1990). Thus, they have less reason to generalize, at least in the short run.

Even with well-practiced strategies, however, it is surprisingly difficult to override usual selection procedures through metacognitive means in order to produce wider generalization of a given approach. This was shown in the domain of children’s arithmetic by an experiment reported in Siegler
Second graders were given subtraction problems under conditions in which they were told that only accuracy, speed, or speed and accuracy were important. The children heeded the instructions; they were fastest and least accurate when told that only speed was important, and slowest and most accurate when told that only accuracy was important. However, the instructions had no effect on their frequency of use of any of the four most common strategies in the experiment. Instead, the children just executed the same strategies more carefully or more quickly, depending on the instructions. Thus, at least in domains such as arithmetic, in which children have substantial experience, strategy choice seems to be a relatively automatic, hard-to-change process. This may underlie the slow generalization of newly taught strategies that has been so frequently observed and lamented (e.g., Brown, Bransford, Ferrera, & Campione, 1983). As noted by Kuhn, Schauble, and Garcia-Mila (1992), learning not to use old strategies may often be as large a challenge as learning to use new ones.

**Individual Differences**

A great deal of research has been devoted to testing relations between individual children's metacognitive knowledge and their memory performance. This research has produced varying results. Early studies often yielded no significant relations between the two (Cavanaugh & Perlmutter, 1982). More recent research, such as the studies by Justice described earlier, has found significant relations. However, these relations also have tended not to be very strong. For example, two meta-analyses of the literature, one weighted by sample size and the other unweighted, revealed identical average correlations of $r = .41$ between individual children's metacognitive knowledge about memory and their recall (Schneider, 1985; Schneider & Pressley, 1989). Thus, the fact that an individual child knows a lot about memory is not extremely helpful in predicting how much the child will remember.

Even these modest to moderate correlations may overstate the relation between long-term metacognitive knowledge and performance. The studies in the meta-analyses included ones in which the assessments of metacognitive knowledge were done after the relevant performance, as well as before. The timing made a big difference: When metamemory was measured after memory performance, the correlations averaged $r = .54$, whereas when it was measured before, the correlations averaged only $r = .25$ (Schneider & Pressley, 1989). The higher correlations produced when metacognitive knowledge was measured after the child performed the task suggests that children use short-term recall of their situation-specific experience to guide their metacognitive judgments, rather than the usual assumption that enduring general metacognitive knowledge governs the judgments.
Evaluation

These first generation models, that emphasized explicit, verbalizable metacognitive knowledge, were clearly insufficient to account for strategy choices. Different investigators drew different conclusions about how to proceed. Some emphasized the need for better assessments of metacognitive knowledge (e.g., Cavanaugh & Perlmutter, 1982). Others argued that because many factors influence strategy choice and cognitive performance, the moderate relations were all that could be expected (e.g., Flavell, 1981; Schneider & Pressley, 1989). Yet others emphasized the need to go on to determine where and when metacognitive knowledge is most strongly related to cognitive activity (e.g., Wellman, 1983).

Another, more radical, conclusion was that enduring, stable metacognitive knowledge has less impact on strategy choices than we think. This interpretation led Flavell (1985) and Brown and Reeve (1986) to suggest that strategy choices may often be generated by unconscious, automatic, nonrational processes, rather than the types of explicit, conscious, rational ones usually thought of as metacognitive. For example, Flavell (1985) wrote:

Your reactions to external or internal memory materials and to your own metacognitive experiences regarding them may often be automatic or reflex-like. Through years of experience as a rememberer (and forgetter), you have learned to recognize and respond adaptively to numerous “patterns” of memory-relevant materials and feelings— and to do so quickly and automatically with little or no conscious reflection. (pp. 234–235)

If this is the case, then abstract, rational knowledge about cognition may not be the place to look for the factors that typically drive strategy choices. Instead, intelligent strategy choices may arise from application of simpler, more basic processes. This was the perspective that motivated development of the second generation strategy choice model.

GENERATION 2: THE DISTRIBUTIONS OF ASSOCIATIONS MODEL

The distributions of associations model was developed by Siegler and Shrager (1984) to show how simple cognitive processes could produce adaptive strategy choices without anything resembling an executive processor. It was specifically aimed at accounting for preschoolers' strategy choices in solving simple addition problems. In this section, we describe the model's basic structure and functioning, the mechanisms that produced
changes in its performance, and the performance and changes that the model produced.

The Model's Basic Structure

As shown in Fig. 2.3, the two main parts of the distributions of associations model are 1) a representation of knowledge about particular problems, and 2) strategies that operate on the representation to produce answers. The answers, in turn, reshape the representation; the model learns by doing.

Within this model, the representation of knowledge is hypothesized to include associations between problems and potential answers, both correct and incorrect. For example, $3 + 5$ would be associated not only with $8$, but also with $6$, $7$, and $9$ (Fig. 2.4). Each problem's associations with various answers can be classified along a dimension of *peakedness*. In a problem with a *peaked distribution*, such as that on the left in Fig. 2.4, most of the associative strength is concentrated in a single answer, ordinarily the correct answer. At the other extreme, in a *flat distribution*, such as that on the right of Fig. 2.4, associative strength is dispersed among several answers, with none of them forming a strong peak.

The process that operates on the representation involves three sequential phases, any one of which can produce an answer: retrieval, elaboration of

---

**FIG. 2.3.** Overview of distributions of associations model.
the representation, and application of an algorithm. In the specific case of preschoolers' addition, children would first try to retrieve an answer. If not sufficiently confident of any answer, they would elaborate the representation of the problem, perhaps by putting up fingers to represent the two addends. If they still did not know the answer, they would use the algorithm of counting the objects in the elaborated representation—in this case, the fingers that were up. All distributions of associations models include these three phases. The way in which retrieval occurs is also constant across the models, though the particulars of the other two phases are specific to each task.

The retrieval mechanism is central. When presented with a problem, children are hypothesized to set a confidence criterion, which determines how sure they must be to state a retrieved answer, and a searchlength, which determines how many attempts they will make to retrieve an answer before trying a different approach to solving the problem. Then the child retrieves an answer. Probability of retrieving any given answer to a problem is proportional to that answer’s associative strength relative to the total associative strength of all answers to the problem. For example, if a given answer had an associative strength of .4, and the total associative strength of all answers was .8, then that answer would be retrieved on 50% of retrieval efforts. This retrieval procedure closely paralleled that hypothesized by Gilliland and Shiffrin (1984).

The retrieved answer is stated if its associative strength exceeds the confidence criterion. For example, if a girl had a distribution of associations and a confidence criterion like that shown for 3 + 5 in Fig. 2.4, she would state the answer if she retrieved 6, 7, or 8, but not if she retrieved any other answer.

If the answer's associative strength does not exceed the confidence
criterion, and the number of retrieval attempts has not exceeded the searchlength, the child again retrieves an answer from the distribution of associations. She states it if its associative strength exceeds the confidence criterion. If the retrieval process fails to yield such an answer within the allocated number of searches, the child elaborates the representation. In the case of simple addition, this occurs through putting up fingers corresponding to the number of objects in each addend or forming a mental image of objects corresponding to that number of objects. A single further retrieval attempt is made, to see if the the kinesthetic and/or visual cues associated with such elaborations allow an answer to be retrieved whose strength exceeds the confidence criterion. If this criterion is met, the answer is stated; if not, the child uses an algorithmic procedure to solve the problem. In the case of addition, this involves counting the number of objects in the elaborated representation.

The distributions of associations model has been formalized within running computer simulations of 4- and 5-year-olds' addition (Siegler & Shrager, 1984), 5- and 6-year-olds' subtraction (Siegler, 1987a), and 8- and 9-year-olds' multiplication (Siegler, 1988b). The simulation of preschoolers' addition is representative. It can be described as follows:

1. The simulation is presented the 25 problems with both addends between 1 and 5, in accord with estimates of their relative frequency obtained through empirical studies of parental presentation of problems (or in the case of older children's arithmetic, through empirical studies of textbook presentation rates).

2. Before each problem, the simulation generates a confidence criterion and a searchlength, whose values vary randomly within the limits set by the simulation.

3. The probability of retrieving an answer is proportional to its associative strength, compared to the associative strengths of all answers to the problem. A retrieved answer is stated if its associative strength exceeds the current confidence criterion. Retrieval attempts continue until either the associative strength of a retrieved answer exceeds the confidence criterion or the number of searches matches the search length.

4. If no answer has been stated and the search length has been reached, the program generates an elaborated representation. The particular elaboration varies with the operation being modeled; in all cases, though, it may lead directly to a statable answer. In the case of addition, the elaboration involves either forming a mental image of counters corresponding to the addends, or putting up one's fingers to represent each addend. The visual and/or kinesthetic cues associated with these elaborations add associative strength to the answer corresponding to the number of fingers that have been put up—usually the correct answer.
5. If no answer has been stated, the model uses an algorithmic backup strategy, which again is specific to the operation being modeled. This algorithmic strategy always yields a statable answer. In the case of addition, the algorithm involves counting the objects in the mental image or the fingers that were put up.

6. Crucial to the overall working of the model is the learning mechanism. Every time the system advances an answer, the association between that answer and the problem increases. The increment is twice as great for correct answers, that presumably are reinforced, as for incorrect answers, that presumably are not. The change in the association value is identical regardless of whether the answer is produced through retrieval or through use of a backup strategy.

The Model's Performance

The model’s functioning can be understood in terms of how it generates the key strategy choice phenomena previously described.

Variability. Variability of strategy use is built into the distributions of associations model. For example, the Siegler and Shrager (1984) simulation of preschoolers' addition included four strategies: retrieval, fingers, counting fingers, and counting. The strategies are particular to preschoolers' addition, and their specifics are less important in the present context than is the organization of strategies and the way in which this organization generates variability. The model considers the strategies in a fixed order in which it first tries retrieval; then, if retrieval is unsuccessful, it tries the fingers strategy (elaboration of the representation); then, if both retrieval and fingers have failed to yield an answer, it uses either the counting fingers strategy or the counting strategy (algorithmic approaches).

This organization yields variability of strategy use within as well as between problems. Each strategy can be, and is, applied to any problem. Siegler and Shrager (1984) found that in the course of the simulation's run, all 4 strategies were applied to each of the 25 problems. This assignment of strategies to problems is not random; problems with peaked distributions of associations elicit greater reliance on retrieval, and problems with flatter distributions elicit greater reliance on the three backup strategies. Nonetheless, strategy use is variable within as well as between problems.

Adaptive strategy choices. At the heart of the model is its procedure for choosing whether to use retrieval or a backup strategy on a problem. The procedure illustrates how adaptive strategy choices can be generated without any homuncular executive processor.

Within the model, adaptive strategy choices between retrieval and backup strategies arise because the peakedness of a given problem's distribution of
associations determines both problem difficulty and the likelihood of using a backup strategy. To understand this view, it is useful to compare the model's workings on problems with peaked and flat distributions of associations.

Relative to a peaked distribution, a flat distribution elicits a higher percentage of use of backup strategies (because flat distributions, by definition, lack a strongly associated answer that has a high probability of being retrieved and a high probability of exceeding the confidence criterion once it is retrieved. The absence of such a strongly associated answer will lead to children often being unable to state any retrieved answer and instead using a backup strategy). The flat distribution will also elicit a higher percentage of errors (because the difference between the strength of association of the correct answer and incorrect ones will be smaller in the flatter distribution, leading to statement of a greater proportion of wrong answers on retrieval trials). Finally, the flatter distribution will lead to longer solution times (because the flatter the distribution, the less likely that an answer whose associative strength exceeds the confidence criterion will be retrieved and stated on an early retrieval attempt). Thus, within this model, backup strategies will be used primarily on the most difficult problems because the peakedness of the distribution of associations determines both problem difficulty and how often retrieval of a stable answer will be possible.

The data on children’s performance supported these predictions and demonstrated the sufficiency of the hypothesized mechanism to generate adaptive choices between retrieval and backup strategies. Within the simulation, correlations between strategy choices (percent use of backup strategies on each problem) and the measures of problem difficulty (percent errors on that problem; length of solution times on each problem) exceeded \( r = .90 \). The simulation’s strategy choices paralleled those of children; the correlation of percent backup strategies on each problem of the simulation and of the children in Siegler and Shrager (1984) exceeded \( r = .80 \). The most difficult problems elicited the highest percentage of backup strategies in both cases. Thus, the simulation generated adaptive strategy choices similar to those of children.

The model also made a specific, nonintuitive prediction regarding the sources of these correlations: The predicted high correlations between percent backup strategies, percent errors, and length of solution times on each problem was really a prediction regarding percent backup strategy use, percent errors on retrieval trials, and length of solution times on retrieval trials on each problem. The reason is that only on retrieval trials do percent errors and length of solution times stem from the peakedness of the distributions of associations. On backup strategy trials, they derive from the difficulty of executing the backup strategies.

Analyses of 4- and 5-year-olds’ performance supported this prediction
(Siegler & Shrager, 1984). Correlations between percent backup strategy use and percent errors on retrieval trials on each problem were significantly higher than correlations between percent backup strategy use and percent errors on backup strategy trials on the problem. The same was true for the corresponding predictions regarding solution times. These nonintuitive predictions arose specifically from the distributions of associations model; it is unlikely that they would have been made without it.

**Change.** The distribution of associations model focused on a single change process: how increasingly peaked distributions of associations lead to faster and more accurate performance and more frequent use of retrieval. This view raises the question of how some problems come to have more peaked distributions than others—that is, how some problems come to elicit higher percentages of errors, longer solution times, and higher percentages of backup strategies than others.

The basic assumption of the model regarding creation of these distributions is that people associate whatever answer they state, correct or incorrect, with the problem on which they state it. This assumption reduces the issue of what factors lead children to develop a particular distribution of associations on each problem to what factors lead them to state particular answers on each problem.

Three factors that seem to lead to differences among problems in peakedness are differences in difficulty of executing backup strategies on the problem, differences in related problems, and differences in frequency of encountering problems. First, consider differences in difficulty of executing backup strategies on each problem. In preschoolers' addition, the most common backup approach is the sum strategy. Children are more likely to correctly execute this backup strategy on problems with small addends, because such problems can be solved via the sum strategy with fewer operations and therefore less chance of error than can other problems. Generating the correct answer via backup strategies on a greater percentage of attempts on these problems leads to their having more peaked distributions of associations.

Intrusions from related operations also influence the rate of acquiring peaked distributions. Knowledge from one numerical operation often intrudes on performance on another; for example, $4 + 3$ fairly often elicits the answer 12, and $4 \times 3$ the answer 7 (e.g., Miller & Paredes, 1990). With preschoolers, knowledge of the counting string often intrudes into addition, leading them to wrong answers (e.g., $3 + 4 = 5$, $3 + 5 = 6$) but also to right ones (e.g., $1 + 2 = 3$, $1 + 3 = 4$). These counting string associations interfere with learning of answers to the first pair of problems, but may facilitate learning of answers to the second pair.

Children also encounter some problems more often than others. For
example, tie problems, such as 2 + 2 and 3 + 3, seem to be presented especially often to preschoolers by their parents (Siegler & Shrager, 1984). More frequent presentation of these problems contributes to their more quickly coming to have peaked distributions.

The effects of these factors—ease of execution of backup strategies, intrusions from related operations, and frequency of problem presentation—have been examined empirically in addition, subtraction, and multiplication. Each of the three factors hypothesized to contribute to learning in these areas has been found to do so. For example, in analyses of preschoolers' addition errors on different problems, each of the three factors added significant independent variance to that which could be accounted for by the other two factors. Together, they accounted for more than 80% of the variance in percentage of errors on the 25 problems (Siegler & Shrager, 1984).

The correlations between the simulation's behavior and that of children increased substantially during its run. Before the learning phase, the correlation between the model's and the children's frequency of use of backup strategies on each problem was $r = -0.03$; the correlation between their solution times was $r = 0.00$. After the learning phase, the corresponding correlations were $r = 0.87$ and $r = 0.80$. Thus, not only did the model's absolute level of performance change in the way that children's did, but the relative performance it generated on different problems also closely resembled that of children.

**Generalization and Individual Differences.** Simulations embodying the distributions of associations model were silent about both generalization and individual differences, but for different reasons. Their silence about generalization was due to their inherently not being able to generalize. All learning was specific to the particular problems that were encountered. Thus, strategy choices on a newly encountered problem would reflect only experience with that problem, regardless of what had happened with previously learned problems.

The silence regarding individual differences was not due to any such conceptual difficulty. Rather, it reflected a lack of implementation within the simulations of ideas regarding individual differences. Such implementations could have been undertaken, but they were not.

**Evaluation**

The distributions of associations model had a number of strengths. It was far more explicit than the previous metacognitive models of strategy choice. It accounted straightforwardly for a number of key phenomena regarding strategy choices and made specific, testable, nonintuitive predictions that
proved to be correct. It illustrated the viability of an alternative to the metacognitive perspective: Rather than adaptive strategy choices implying a knowledgeable and insightful executive processor, the choices could arise through the operation of simple cognitive processes such as retrieval.

However, the model had certain weaknesses as well. It was too inflexible, too limited in its explicitness, and too dumb.

**Inflexibility.** The distributions of associations model was rigid. It always considered strategies in the same order, regardless of the circumstances. It also did not provide any obvious way in which new strategies could be integrated into the strategy choice process.

Both properties are at odds with what is known about human strategy choices. People do not always attempt retrieval before other strategies. Instead, they at times first consider the strategy that seems the most likely to pay off, even when that strategy is not retrieval (Reeder, 1982). Further, adults under time pressure often choose to use calculational strategies in less than the amount of time that retrieval requires (Reeder & Ritter, 1992). People's strategy choice procedures clearly are more flexible than those described within the distributions of associations model.

Another way in which the distribution of associations model was inflexible involved the fixed three phase process of retrieval, elaboration of the representation, and use of a solution algorithm. This three phase approach fit the particular strategies used in preschoolers' addition (and in somewhat older children's beginning subtraction and multiplication, as well), but seemed too restrictive to capture strategy choices in general. Even within basic addition, the min strategy does not fit in any natural way into the three-phase structure. Thus, the goal of generating a more general model of strategy choice demanded a more flexible structure.

**Limited Explicitness.** The distribution of associations model was far more explicit than previous metacognitively oriented models of how strategy choices are made. However, the explicitness lay primarily in the depiction of the choice between stating a retrieved answer or using a backup strategy. Procedures for choosing among alternative backup strategies were left vague. For example, the choice between elaborating the representation by putting up fingers or by forming an internal representation was simply stated as a pair of probabilities. There was no account of how this decision was made.

**Dumbness.** The distributions of associations model was considerably less intelligent than children are in at least two ways. First, it could not draw any generalizations from its experience. No matter how much experience it had with solving problems, it could not draw any implications regarding
other problems, even the most closely related. Its learning was all literal. It would not even draw the generalization that a strategy that was useful on 5 + 3 might also be useful on 3 + 5. Yet, even infants generalize problem solving approaches to related problems (Rovee-Collier, 1989).

The model also had no abstract knowledge about the usefulness of strategies or the difficulty of problems. As described earlier, children’s judgments of problem difficulty correlated about \( r = .50 \) with the actual difficulty of the problems (Siegler & Shrager, 1984). This is insufficient to account for their very adaptive strategy choices, but it also is not negligible. The model did not include any data that would provide a basis for such judgments.

In response to these limitations in flexibility, range of explicit depictions of strategy choice mechanisms, and breadth of knowledge, we recently built a new simulation—the Adaptive Strategy Choice Model (ASCM, pronounced “Ask-em”). The goal was to create a more flexible, more precise, and more intelligent model of strategy choice. We also wanted to simulate acquisition of knowledge not just of small addend problems in the preschool years, but of all single-integer problems over the preschool and elementary school period. Thus, the simulation is intended to depict the mastery of single-digit addition over the period from roughly 4 to 12 years of age.

**GENERATION 3: ASCM**

The Model’s Basic Structure

Fig. 2.5 illustrates ASCM’s overall organization. Strategies operate on problems to generate answers. The solution process yields information not only about the answer to the particular problem, but also about the time required to solve the problem using that strategy and the accuracy of the strategy in answering the problem. This information is used to modify the database regarding the strategy, the problem, and their interaction.

*The Database.* The type of information that gets entered into the database is illustrated in Fig. 2.6. Through their experience solving problems, children gain knowledge of both strategies and problems. Knowledge of each strategy can be divided into knowledge based on actual data and knowledge based on projections (inferences) from that data.

The actual data include each strategy’s past speed and accuracy aggregated over all problems (*global data*), its speed and accuracy on problems with a particular feature (*feature data*), its speed and accuracy on each particular problem (*problem-specific data*), and its newness (*novelty data*).
The roles of the first three types of data should be easy to comprehend, but that of the novelty data may require some explanation. Inclusion of these data was motivated by an attempt to answer the question: How do new strategies come to be used in situations where existing strategies work well? In the context of young children's addition, if a child can consistently solve a problem by using the sum strategy, why would the child ever try the min strategy on the problem? ASCM deals with this issue by assigning novelty points to newly discovered strategies. These novelty points temporarily add to the strength of new strategies, and thus allow the new strategies to be tried even when they have little or no track record. With each use of a new strategy, some of its novelty strength is lost, but information about its speed and accuracy is gained. This leads to the strategy's probability of use increasingly being determined by the expanding database on its effectiveness. The idea of novelty being a kind of strength was suggested by the observation that people (especially children) are often interested in exercising newly acquired cognitive capabilities (Piaget, 1970) and by the realization that without a track record, a newly acquired strategy might otherwise never be chosen, especially if reasonably effective alternatives were available.

Whenever ASCM is presented a problem, it uses these speed, accuracy, and novelty data for each strategy to make projections concerning how well
the strategy is likely to do in solving the problem. If a strategy has never been used on a particular problem, ASCM's projections are based solely on global and featural data. If the strategy has never been used on the particular problem or on any problem with that feature, only global data are used to derive the projection.

The Model's Operation. ASCM is implemented as a running computer simulation. Because this simulation has not been described previously, its operating assumptions are explained here in some detail.

1. At the beginning of its run, ASCM knows only the two strategies that are common among 4-year-olds—retrieval and the sum strategy—and basic procedures for choosing strategies, collecting data on the outcomes they generate, and projecting their future usefulness. These latter competencies are hypothesized to be basic properties of the human information processing system and to be present from birth.

2. During the learning phase, the simulation is repeatedly presented the 81 basic addition facts formed by all possible combinations of addend values 1–9, inclusive. The problems are presented equally often. In the absence of data on presentation rates over the large age range being modeled, this seemed the most conservative assumption.
3. After a number of exposures to each problem (60 trials/problem in the simulation runs reported here), the min strategy is added to those initially available. This is done to correspond to the time, usually sometime during first grade, when children add the min strategy to their repertoire. The process of discovery of the min strategy is not yet modeled. Like its predecessor, the new simulation focuses on choices among existing strategies, rather than on how new strategies are constructed.

4. Strategy choices are based on the projected strength of each strategy. As shown in Table 2.3, projected strength is a function of the strategy's past speed and accuracy on problems as a whole, on problems with features in common with the current one, and on the particular problem being solved. For new strategies, the strategy's novelty boosts its strength beyond what its past performance alone would justify.

5. A logistic equation weights these sources of data according to the amount of information they reflect. When a strategy has rarely been used on a particular problem, global and featural data are weighted most heavily. As more information becomes available about how well the strategy works on the problem, problem-specific information receives increasing weight, eventually exercising the largest influence. The reasoning is that data derived from a few uses of a strategy is inherently noisy, but problem-specific information based on a substantial database is the best predictor of a strategy's future effectiveness on that problem. A similar logistic equation is used to weight data according to how recently it was generated. The reasoning underlying this decision was similar: Recent performance is given greater weight because it is likely to better predict future performance.

6. Each time a problem is presented, these weighted sources of information provide the input to a stepwise regression equation that computes the projected strength of each strategy on the problem.

7. Probability of choosing a particular strategy is proportional to that strategy's projected strength relative to that of all strategies combined. The simulation attempts to execute whichever strategy is chosen. If a backup strategy is chosen, it is executed to completion. If retrieval is chosen, a

<table>
<thead>
<tr>
<th>Table 2.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Equations Governing ASCM's Operation</td>
</tr>
<tr>
<td>Strength (strategy) = f (speed, accuracy, and novelty)</td>
</tr>
<tr>
<td>P(retrieve strategy) = Strength (strategy) / Strength of all strategies</td>
</tr>
<tr>
<td>P(retrieve answer) = Strength (answer) / Strength of all answers</td>
</tr>
</tbody>
</table>
procedure identical to that within the distribution of associations model is followed. This means that when an answer is retrieved with associative strength that exceeds the confidence criterion, the answer is stated.

8. If no storable answer is retrieved and the searchlength has been reached, the model returns to the strategy choice phase and chooses among the backup strategies. The process is the same as at the beginning of the trial, except for the exclusion of retrieval from the set of strategies under consideration (because it has already been tried). Thus, the probability of a given backup strategy being chosen at this point reflects its projected strength relative to that of all backup strategies combined.

9. Probabilities of errors using the sum and min strategies are proportional to the number of counts required to execute the strategy on that problem. The errors arise through double counting or skipping an object in the representation of the problem. Each count entails a probability of error; thus, the greater the number of counts, the more likely that an error will be made. On retrieval trials, errors arise through an incorrect answer being retrieved and having sufficient associative strength to be stated.

10. Solution times on backup strategy trials are proportional to a constant times the number of counts that are executed. The constant is smaller for the sum than for the min strategy, because children take less time per count in counting from one than in counting from other numbers (Siegler, 1987b). Times on retrieval trials reflect a constant times the number of searches prior to locating a storable answer. This constant is much smaller than those used with the backup strategies, reflecting the fact that retrieval is much faster than the sum or min strategies.

11. As in the distributions of associations model, each time an answer is advanced, ASCM increases the association between that answer and the problem. Unlike in the earlier model, ASCM also adds information regarding the speed and accuracy with which the answer was generated to the database for the strategy.

12. Each execution of a backup strategy also brings an increase in the strategy’s speed and a decrease in its probability of generating an error. Thus, strategy execution improves with practice.

There are clear similarities between ASCM and the distributions of associations model; in a sense, ASCM is a generalized version of the principles inherent in its predecessor. Several differences between the two should also be noted, though. ASCM is smarter; it possesses mechanisms for generalizing its experience to unfamiliar problems. It also is explicit about a broader range of strategy choices, in that it describes how choices among alternative backup strategies are made. Further, it is more flexible in allowing strategies to be considered in any order.
The Model's Performance

As in the tests of the distributions of associations model, the tests of ASCM's performance involved a learning phase of a given length and then a test phase that indicated the performance generated by the simulation after that amount of experience. During the learning phase, performance on each trial altered the database regarding strategies, problems, and answers. The analogy was to experience children would have had prior to entering the experimental situation. During the test phase, in contrast, the database remained constant. The analogy was to children's performance in the experimental situation, after their having had a given amount of pre-experimental experience.

The learning phases ranged from 60 to 1,250 trials per problem. This latter figure may at first sound high. However, to put it in perspective, 1,250 trials per problem works out to approximately 100 simple addition operations per school day over 6 years of elementary school (6 years, 180 school days/year, 81 problems). Given the torrent of addition problems that elementary school students receive from textbooks, workbooks, handouts, fact quizzes, and magic minutes, plus the embedding of simple addition within the multidigit addition and multiplication algorithms (a typical 3-digit by 3-digit multiplication problem entails 16 adding operations), this number does not seem unreasonable. Where not otherwise specified, the values reported for the simulation will be those attained after a learning phase of 750 exposures per problem, a point at which performance is very good, but where ceiling effects are not a serious difficulty.

Regardless of the length of the learning phase, the test phase included 3,700 trials per problem (300,000 total trials). Because the simulation did not undergo any change during the test phase, the only effect of this large number of trials was to approximate closely the expected values the simulation would produce after that length learning phase. The other key parameter values within the simulation are shown in Table 2.4.

<table>
<thead>
<tr>
<th>TABLE 2.4</th>
<th>ASCM’s Main Parameter Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning phase length: 60, 120, 250, 500, 750, or 1,250 trials/proble</td>
<td></td>
</tr>
<tr>
<td>Test phase length: 3,700 trials/problem</td>
<td></td>
</tr>
<tr>
<td>Range of confidence criteria: .01–.95</td>
<td></td>
</tr>
<tr>
<td>Range of search lengths: 1–3</td>
<td></td>
</tr>
<tr>
<td>Timing of min strategy introduction: After 60 trials/problem</td>
<td></td>
</tr>
<tr>
<td>Range of min strategy percent correct: 30%–95%</td>
<td></td>
</tr>
<tr>
<td>Range of sum strategy percent correct: 0%–85%</td>
<td></td>
</tr>
<tr>
<td>Increment in associative strength for statement of correct answer: .0002</td>
<td></td>
</tr>
<tr>
<td>Increment in associative strength for statement of incorrect answer: .0001</td>
<td></td>
</tr>
</tbody>
</table>
As with the distributions of associations model, ASCM can be understood in terms of its treatment of variability, strategy choice, change, generalization, and individual differences. Its performance on these dimensions highlights both its similarities to and its differences from its predecessor.

Variability. ASCM generates variability in the senses that the earlier model did, and in some additional senses as well. Like the distributions of associations model, it uses diverse strategies both within and between problems. It tends to most often use strategies on the problems where they work best, but strategy use varies within as well as between problems.

ASCM also generates some types of variability that the distributions of associations model did not. In the distributions of associations model, strategies were always executed in a fixed order, with retrieval invariably being tried first. Within ASCM, in contrast, any strategy can be tried first. In practice, retrieval comes to be tried first in the large majority of cases (99% after a learning phase of 750 trials/problem). This is due to its always being the fastest strategy and usually being accurate when it is used. However, the same mechanism that allows the novelty points to add associative strength to a strategy also opens ASCM to situational influences that can temporarily boost the strength of competing strategies (e.g., through instructions encouraging their use or through a child consciously thinking that it would be a good idea to try a particular strategy). Further, the order in which the backup strategies are considered varies in practice as well as in theory from trial-to-trial, unlike in the distributions of associations model, where their order was fixed.

Adaptive Strategy Choices. ASCM produced performance that was adaptive in the same ways as the performance generated by the distributions of associations model, and in an additional sense as well. Like the earlier model, it produced extremely high internal correlations between each problem's percent use of backup strategies and percent errors ($r = .98$) and between each problem's percent use of backup strategies and its mean solution time ($r = .99$).

We also compared ASCM's performance to that of children. The children were 120 students at a middle class suburban school, tested near the end of first grade. Speed, accuracy, and strategy use on each trial were assessed as in previous studies (e.g., Siegler, 1987b; 1989b). For each problem, percent correct, median solution time, and percent use of retrieval was computed over the set of 120 children.

As shown in Table 2.5, the simulation's percentages of use of backup strategies on each problem initially were almost uncorrelated with those of the children. However, as the simulation gained experience, the patterns of
TABLE 2.5
Correlations Between ASCM's and Children's Performance on Each Problem

<table>
<thead>
<tr>
<th>Measure</th>
<th>60</th>
<th>250</th>
<th>500</th>
<th>750</th>
<th>1,250</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent errors</td>
<td>.27</td>
<td>.78</td>
<td>.85</td>
<td>.85</td>
<td>.50</td>
</tr>
<tr>
<td>Mean solution times</td>
<td>.12</td>
<td>.55</td>
<td>.77</td>
<td>.90</td>
<td>.81</td>
</tr>
<tr>
<td>Percent retrieval</td>
<td>.06</td>
<td>.21</td>
<td>.76</td>
<td>.93</td>
<td>.78</td>
</tr>
</tbody>
</table>

Performance grew increasingly similar. After 750 trials per problem, ASCM's percentages of use of backup strategies on each problem correlated \( r = .93 \) with those of the children. After this point, the correlations decreased, for a reason that is easy to understand. After a learning phase with 1,250 trials per problem, the percentage of test-phase errors made by the simulation was extremely low (1%) and the percentage of strategies other than retrieval was also quite low (5%). This led to there being too little variance to produce the high correlations seen after shorter learning phases. Nonetheless, even in this advanced state, the simulations still produced adaptive choices between use of a backup strategy or retrieval much like those of children.

This type of adaptive strategy choice was also made by the distributions of associations model. However, ASCM also made a second type of adaptive choice, the choice of which backup strategy to use, that the earlier model did not. The adaptiveness of ASCM's choices is evident in the frequency of use of the min strategy relative to that of both backup strategies (percent min strategy/percent min + sum strategies). After relatively brief learning phases (e.g., 120 trials per problem), the best predictor of min strategy use on a problem was the size of its smaller addend; the smaller this value, the higher the percentage of min strategy use on the problem. After greater numbers of trials, the best predictor of percentage of min strategy use on a problem was the difference between the problem's addends; the larger the difference, the greater the percentage of min strategy use. These are the same variables that best predict min strategy use (given that retrieval was not used) in empirical data on children (Siegler, 1987b). It also makes intuitive sense for the min strategy to be used most often on problems with small smaller addends, where it is easiest to execute correctly, and on problems with large differences between the addends, where the min strategy produces the greatest reduction in counting relative to the sum strategy.

Change. Perhaps the single most essential property of a simulation of acquisition of arithmetic knowledge is that it progress from the inaccurate performance characteristic of children just beginning to add, to the virtually
perfect performance that characterizes children by fifth or sixth grade. ASCM met this test. After a learning phase of 60 trials per problem, its performance was 31% correct, whereas after a learning phase of 1,250 trials per problem, its performance was 99% correct. As shown in Fig. 2.7, this improvement did not occur in sudden jumps; rather it came steadily with a substantial amount of experience. Children show similar gradual improvement with experience in the accuracy of their addition (Kaye, Post, Hall, & Dineen, 1986; Siegler, 1987b).

As shown in Fig. 2.8, ASCM’s changes in frequency of use of the three strategies also paralleled children’s. At first the simulation used only the

**Learning Trials / Problem**

FIG. 2.7. ASCM’s percent correct after different length learning phases.
sum strategy and retrieval (the only strategies it knew), with the sum strategy being employed on the large majority of trials. After the min strategy was added, it became the most frequently used strategy, with the sum strategy and retrieval also being used on substantial numbers of trials. This corresponds to children's performance at around first grade. Beyond this point, use of both sum and min strategies decreased, and retrieval became increasingly dominant. After a learning phase of 1,250 trials per problem, retrieval was used on 95% of trials.

A third key type of progress involved decrements in solution times. The overall decrement in times was produced by two factors: shifts from the slower backup strategies to the faster retrieval approach, and faster execution of each strategy. Along with this general speedup, ASCM produced changes in the pattern of solution times on different problems,
changes that paralleled those in the pattern of children's times on different problems. Empirical studies of children's arithmetic have demonstrated that the best predictor of children's solution times on addition problems changes with age in a regular way. In the preschool period, the size of the sum is the best predictor (Siegler & Robinson, 1982; Siegler & Shragen, 1984); in first and second grade, the size of the smaller addend is the best predictor (Groen & Parkman, 1972; Groen & Resnick, 1977; Svenson & Broquist, 1975). Beyond this time, the product of the two addends tends to be most predictive (Geary, Widaman, & Little, 1986; Geary, Widaman, Little, & Cormier, 1987; Miller, Perlmutter, & Keating, 1984). Accounting for these changes in the best predictors of solution times is a considerably more rigorous test of the simulation than simply producing improvements in the absolute levels of the times.

As shown in Table 2.6, the simulation underwent the same type of changes in the predictors of its solution times as did children. At first, the best predictor of solution times was the sum of the addends; later, the size of the smaller addend was the best predictor; still later, product size became the most predictive variable.

The reasons for these predictive relations also appear to be the same for the simulation and the children. The sum was the best predictor at first because the predominant strategy that was used, the sum strategy, generated solution times that are a linear function of the sum (Siegler, 1987b). Smaller addend size became the best predictor of solution times when the min strategy was most often used, because it generated solution times that are a linear function of the size of the smaller addend (Siegler, 1987b).

The reasons why the product was the best predictor of both the children's and the simulation's times are both less intuitive and more interesting. The only previous explanation of the phenomenon was Widaman, Geary, Cormier, and Little's (1989) spreading activation model. Within this model, arithmetic knowledge is represented much like the type of facts table often seen in elementary school textbooks. In such tables, the numbers 0–9 head

<table>
<thead>
<tr>
<th>Learning Trials per Problem</th>
<th>Best Predictor</th>
<th>% Variance Accounted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>Sum</td>
<td>91</td>
</tr>
<tr>
<td>120</td>
<td>Smaller addend</td>
<td>88</td>
</tr>
<tr>
<td>250</td>
<td>Product</td>
<td>93</td>
</tr>
<tr>
<td>500</td>
<td>Product</td>
<td>91</td>
</tr>
<tr>
<td>750</td>
<td>Product</td>
<td>79</td>
</tr>
<tr>
<td>1,000</td>
<td>Product</td>
<td>93</td>
</tr>
<tr>
<td>1,250</td>
<td>Product</td>
<td>92</td>
</tr>
</tbody>
</table>
the columns and the rows and the sum of the two addends at each intersection, so that 0 is in the top left hand corner and 18 is in the bottom right hand corner. Geary and Widaman hypothesized that presentation of an addition problem $X + Y$ causes activation to spread from the origin to cover a rectangle whose vertices are 0, 0; $X, 0; 0, Y$; and $X, Y$. The time to activate this rectangle was hypothesized to be proportional to its area, thus leading to the product $XY$ determining solution times to $X + Y$.

Such fact table explanations were faulted on a number of grounds (e.g., Ashcraft, 1987; Pellegrino & Goldman, 1989; Siegler, 1988b). Among the criticisms, the fact-table model renders problematic why factors other than problem size (e.g., presence of 5 as an addend or multiplicand) should make problems easier than their size would suggest. It also makes problematic how people would ever retrieve answers that are not part of the arithmetic table (e.g., $7 \times 8 = 56$).

ASCM generated the same result (solution times being best predicted by the product of the addends) through an entirely different, and we believe more plausible, mechanism. Within ASCM, as within the distributions of associations model, relative difficulty of problems reflects the relative difficulty of executing backup strategies on them. In the case of addition, the two most common backup strategies are the sum and min strategies. Solution times on retrieval trials ultimately reflect the distribution of associations generated by the backup strategies. Because min and sum are the main backup strategies used, solution times should reflect the history of the simulation having sometimes solved the problems using one approach and sometimes the other.

To demonstrate the sufficiency of this explanation to account for the phenomenon, we generated several artificial data sets. In one, we had the simulation use both the min strategy and the sum strategy on exactly half of trials on each problem. Within this run, the solution time predicted for each problem was an unweighted average of the time that would arise from use of each strategy.

The results illustrated that such a history would lead to the product being an excellent predictor of solution times in absolute terms, accounting for 98% of the variance in times on the 81 single-digit problems, and also to the product being more predictive than the other variables that were examined: the smaller addend, larger addend, first addend, second addend, sum, and sum squared.

This result was not dependent on the frequency of strategy use being equal for the two backup strategies. The product was just as good a predictor in absolute terms, and the best predictor in relative terms, on a test in which the min strategy was used on two-thirds of trials on each problem and on a test in which the sum strategy was used on two-thirds of trials on each problem. These results suggest that the product is the best
2. VARIATION, SELECTION, AND CHANGE

predictor of both the children's and the simulation's solution times because it reflects the combined influences of the most frequent backup strategies in shaping strengths of associative connections between problems and answers. The results hold over a range of assumptions about the history of learning that shaped the associative network.

This example illustrates how simulations can suggest alternative explanations of empirical phenomena that would not otherwise be generated. The simulation was not designed with the intention of explaining why the product was the best predictor of older children's and adults' solution times. When the pattern emerged, however, we recognized it as the same one that had emerged in the empirical literature. The result allowed us to say with certainty that such a pattern could arise in the absence of a tabular representation; we know for a fact that ASCM does not have such a representation. The simulation thus served as a tool analogous to animal models in medicine—a system believed to work similarly along the relevant dimensions, but one amenable to much more precise experimental manipulation than the system being modeled. This function seems to us to be one of the most valuable that simulations can serve in the study of cognitive development.

Generalization. As noted earlier, a key requirement for a realistic model of arithmetic is that it be able to generalize its knowledge of strategies to new problems. A first or second grader who, for the first time, needed to mentally calculate \( 4 + 38 \) would be unlikely to know the answer, but almost certainly would choose to count from 38 rather than from 1. Inability to generalize in this way was one of the major limits of the distributions of associations model.

To test ASCM's ability to generalize, we presented it with 50 learning-phase exposures to each of 10 single-digit problems, and then, during the test phase, examined its strategy choices on the other 71 problems. We then repeated the procedure with 500 learning-phase trials per problem for each of the 10 problems, to see if generalization improved with experience. The 10 problems were generated randomly, subject to the stipulation that each integer was included once as the first addend and once as the second addend. We chose to have a low ratio of example problems to generalization problems to mimic the situation that children encounter, in which there are very large numbers of problems that they occasionally need to solve (e.g., \( 3 + 29 \)).

In this test of ASCM's ability to generalize, we were interested in whether it would choose the min strategy most often on problems where it was easiest to execute and/or where its advantage in reduced counting over the sum strategy was greatest. For example, would it choose the min strategy especially often on \( 9 + 1 \), a problem that has both of these properties?
ASCM showed exactly this pattern of generalization. On the 71 problems that had not been presented in the learning phase, the best predictor of percentage of trials on which the min strategy was used was the difference between the addends (Table 2.7). What this meant can be illustrated by considering the use of the min and sum strategies on two specific problems: $9 + 8$ and $9 + 1$. On the former problem, where the difference between addends was only 1, the min strategy was used on 33% of trials during the generalization phase and the sum strategy on 67% (after the simulation had been given 500 exposures to each of the 10 original problems). In contrast, on $9 + 1$, where the difference between the addends was 8, the min strategy was used on 76% and the sum strategy on only 24% of trials.

The differentiation between problems on which the min strategy was more and less helpful increased with learning; differences between the addends was a better predictor of amount of generalization of the min strategy after a learning phase of 500 trials per problem than after one of 50 trials per problem. This made sense, because the model was obtaining increasingly valid data about each strategy's effectiveness on different problems. Even at the first point of measurement, however, ASCM produced generalizations similar to those shown by children.

**Individual Differences.** As noted earlier, Siegler (1988a) found substantial individual differences in children's approaches to addition, subtraction, and word identification tasks. The children could be classified into three distinct groups: the good students, the not-so-good students, and the perfectionists. As the names suggest, the good students were both more accurate and more likely to retrieve answers than the not-so-good students. The perfectionists were as accurate as the good students, but used retrieval even less often than the not-so-good students.

Both the distributions of associations model and ASCM suggested a simple means through which the three individual difference groups could arise: parametric variation in peakedness of the distributions of associations and in confidence criteria. The more peaked the distribution of associations, the more likely that a correct answer will be retrieved, and the more likely that if retrieved, its associative strength will exceed the confidence criterion and be stated. The higher the confidence criteria, the greater the

<table>
<thead>
<tr>
<th>Learning Phase</th>
<th>Best Predictor of Min Strategy Use on Generalization Problems</th>
<th>% Variance Accounted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>Difference between addends</td>
<td>62</td>
</tr>
<tr>
<td>500</td>
<td>Difference between addends</td>
<td>83</td>
</tr>
</tbody>
</table>
associative strength of the retrieved answer must be to exceed it, and therefore, the fewer retrieved answers will be stated.

This view suggests an interpretation of each group's pattern of performance. The good student pattern would arise from a combination of peaked distributions of associations and a wide range of confidence criteria. This would lead to both high accuracy and frequent use of retrieval, because the highly peaked distributions would result in frequent retrieval and statement of the correct answer. The not-so-good student pattern would arise from flat distributions of associations and low confidence criteria. This would generate inaccurate performance and medium-to-low amounts of retrieval, because incorrect answers would often be retrieved and would sometimes exceed the low confidence criteria. The perfectionist pattern would arise from peaked distributions and very high confidence criteria. This would lead to accurate performance, but to low amounts of retrieval, because only the most peaked distributions would have correct answers with enough associative strength to exceed the very high confidence criteria.

To test this interpretation, we created three variants of the simulation. They differed only in their values of the two parameters hypothesized to underlie the individual differences: probability of correct execution of backup strategies (which influenced the peakedness of distributions that were formed) and range of confidence criteria. The simulation of the not-so-good students' performance executed backup strategies less accurately than did the simulations of the perfectionists' and good students' performance, which were identical to each other in accuracy of execution of these strategies. The confidence criteria of the not-so-good students were consistently low (.10-.50), those of the perfectionists consistently high (.50-.90), and those of the good students included both low and high values (.10-.90). Other than these two parameter values, the simulations of the three groups were identical.

Results indicated that the variations in these two parametric values were sufficient to account for the observed pattern of individual differences. As shown in Table 2.8, ASCM's simulation of the not-so-good students produced lower percentages correct than its simulations of the good students and perfectionists, which did not differ. As with children, ASCM's good student simulation produced the greatest amount of retrieval, its not-so-good student simulation the next most, and its perfectionist simula-

<table>
<thead>
<tr>
<th>Group</th>
<th>% Correct</th>
<th>% Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good students</td>
<td>96</td>
<td>84</td>
</tr>
<tr>
<td>Not-so-good students</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Perfectionists</td>
<td>96</td>
<td>89</td>
</tr>
</tbody>
</table>
tion the least. The simulations thus illustrate how qualitatively different patterns of performance can arise through parametric variations within the same basic processing framework.

CONCLUSIONS

The central thesis of this chapter is that transcending the typical 1:1 equation between age and way of thinking, and recognizing the importance of variation and selection in cognitive development, will lead to a better understanding of change. To support this thesis, we identified a set of key phenomena regarding children's strategy choices and examined how three generations of models accounted for them. The increasing ability of the models to account for change reflects two types of progress: 1) Broadening of the types of variability and selection that are accounted for, and 2) Increasingly precise specification of how the variability and selection are produced.

The first-generation metacognitive models represented an initial effort in this direction. They reflected the recognition that individual children often know multiple strategies and that they use them under some circumstances and not others. They depicted choices among the strategies in terms of an executive processor using explicit, conscious knowledge about cognitive capacities, strategies, and situational variables to decide which strategy to use. Change in this knowledge was what led to better choices among the strategies. The key insight of these first generation models was that choices among strategies was a key issue to be explained. The key weaknesses were their questionable assumption that strategy choices are in general produced by explicit, rational analyses of situations, capacities, and strategies, and their lack of specificity regarding how the choice process works.

The second-generation attempt to explain strategy choices—the distributions of associations model—recognized a broader range of variability: variability not just in strategies known to each child but also in answers associated with the problem and in individual children's strategy choices. Within this model, choosing whether to use a backup strategy or retrieval was viewed as reflecting an interaction between the organization of the strategies and the peakedness of the distributions of associations for the particular problem. Change in the peakedness of the distributions of associations changed the frequency with which each strategy was selected, as well as the speed and accuracy of performance. The key insight embodied within the distributions of associations model was that intelligent strategy choices did not require an intelligent executive processor. The model was also far more explicit than the metacognitive models had been about how strategy choices are made and about the factors that contribute to changes
2. VARIATION, SELECTION, AND CHANGE

in the choices. However, it was too inflexible, not explicit about how choices among backup strategies are made, and incapable of generalizing its strategy choices to new problems.

The third generation model, ASCM, further expanded the range of variability that was recognized. It not only attempted to account for variability in strategies, answers, and individual patterns of performance, but also for variability in the order in which strategies were considered. Its choices among strategies made use of a considerably broader range of data—not just associations between problems and answers, but also global, featural, local, and novelty data about the strategies. These data allowed ASCM to make adaptive choices on novel as well as familiar problems, and to choose between alternative backup strategies as well as between retrieval and some backup strategies. Relative to its predecessors, ASCM also specified a larger set of contributors to change in strategy choices—not just changes in the peakedness of the distributions of associations, but also changing knowledge about all of the types of data regarding strategies that contribute to adaptive choices at any one time. The key insight was that the principles used by the distributions of associations model to choose between retrieval and a backup strategy could be used to choose among any set of strategies, and provided a basis for generalization as well. However, ASCM had its limits: Additional sources of variation and selection that need to be explained.

An important source of variation that remains to be accounted for is strategy acquisition. ASCM provides a general account of how choices among existing strategies are made, but does not indicate how the strategies come to be present. Yet, acquisition of new strategies is what makes strategy choice both possible and necessary.

This limit may soon be overcome. Jones and VanLehn (1991) formulated a simulation, GIPS, that models the data on preschoolers' discovery of the min strategy reported by Siegler and Jenkins (1989). If the models can be integrated, GIPS may provide a kind of "front end" to ASCM, providing it with the strategies that it would then choose among.

An important source of selection that remains to be accounted for is explicit, conscious, metacognitive judgment. This may not play a large role when people have had experience in a domain, as is the case in children's arithmetic. However, this type of knowledge seems likely to be more influential when children are presented with a novel problem and need to decide what to do. Metacognitive models may have overestimated the influence of conscious, explicit knowledge, but this does not mean that such knowledge has no influence. Understanding when and how metacognitive knowledge is used to choose among alternative strategies seems essential for a general understanding of strategy choice.

One final comment on a basic assumption underlying the present research
effort seems essential. Science progresses not only through increasingly precise description of an increasing range of phenomena, but also through repeated efforts to explain the phenomena. We believe that models are critical in this effort, because they indicate the range of phenomena that can be accounted for at any one time, and thus provide a benchmark against which the successes and failures of alternative explanations can be measured. Such models never have the solidity of facts and are never complete. Still, they seem to us extremely valuable, because they make clear both what we have achieved and the much larger tasks that remain.

ACKNOWLEDGMENTS

This research was supported by a grant from the Mellon Foundation, by a grant from the Spencer Foundation, and by grant HD-19011 from the National Institutes of Health.

REFERENCES


2. VARIATION, SELECTION, AND CHANGE


2. VARIATION, SELECTION, AND CHANGE


APPENDIX A:
Associative Strength Values for 0-9 + 0-9

<table>
<thead>
<tr>
<th>Second Addend</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>53</td>
<td>53</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>53</td>
<td>51</td>
</tr>
<tr>
<td>1</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>52</td>
<td>50</td>
<td>53</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>53</td>
<td>51</td>
<td>52</td>
<td>47</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>45</td>
<td>42</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>52</td>
<td>53</td>
<td>52</td>
<td>53</td>
<td>46</td>
<td>46</td>
<td>46</td>
<td>42</td>
<td>45</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>53</td>
<td>50</td>
<td>45</td>
<td>47</td>
<td>53</td>
<td>45</td>
<td>38</td>
<td>31</td>
<td>31</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>53</td>
<td>52</td>
<td>48</td>
<td>40</td>
<td>40</td>
<td>46</td>
<td>35</td>
<td>24</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>53</td>
<td>52</td>
<td>44</td>
<td>44</td>
<td>36</td>
<td>37</td>
<td>49</td>
<td>26</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>53</td>
<td>51</td>
<td>48</td>
<td>39</td>
<td>23</td>
<td>26</td>
<td>19</td>
<td>51</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>52</td>
<td>50</td>
<td>46</td>
<td>41</td>
<td>28</td>
<td>12</td>
<td>7</td>
<td>18</td>
<td>38</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>53</td>
<td>52</td>
<td>43</td>
<td>45</td>
<td>24</td>
<td>18</td>
<td>11</td>
<td>8</td>
<td>24</td>
<td>40</td>
</tr>
</tbody>
</table>