MECHANISMS OF COGNITIVE DEVELOPMENT

Robert S. Siegler
Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 15213

CONTENTS

NEURAL MECHANISMS ................................................................. 355
  Synaptogenesis ................................................................. 355
  Segregation of Neural Input ............................................... 356
  Experience-Expectant and Experience-Dependent Processes ....... 357
ASSOCIATIVE COMPETITION .................................................... 359
  Connectionist Models ......................................................... 360
  Associative Competition and Language Development ............. 361
ENCODING ................................................................................. 362
  Recent Empirical Findings .................................................. 363
  How New Encodings Are Formed ........................................... 364
ANALOGY ................................................................................. 365
  Recent Empirical Findings .................................................. 365
  How Analogies Are Generated .............................................. 366
STRATEGY CHOICE ................................................................... 369
  Recent Empirical Findings .................................................. 369
  How Strategy Choices Change ............................................. 372
CONCLUSIONS ........................................................................... 374

To develop is to change. Without a good understanding of the mechanisms that produce change, no comprehensive understanding of development is possible. The difficulty of identifying such mechanisms is well known and has often been noted (e.g. Flavell 1984; Kuhn 1984; Miller 1983). Not as well known or as often noted, however, is the substantial progress that has been made in the past few years in establishing the effects that particular mechanisms produce and in specifying how the mechanisms operate.

Here I review some of the most promising recent ideas about change mechanisms that contribute to cognitive development. I hope that bringing together in a single context a number of these relatively well-worked-out examples will facilitate further advances in theorizing about change mechanisms.
The progress of developmental theories depends critically on improving our grasp of how change occurs (Flavell 1984; Klahr 1984; MacWhinney 1987; Miller 1983; Siegler 1983; Sternberg 1984). Well-specified change mechanisms can increase the generality of theories by revealing that seemingly unrelated acquisitions are products of the same process. They also can help explain within a single framework how early concepts can be acquired quickly while advanced forms of the same concepts can take years to attain. Improved understanding of change mechanisms also offers accurate prediction of nonintuitive empirical phenomena.

What do we mean by a “mechanism of cognitive development”? Broadly defined, a cognitive-developmental mechanism is any mental process that improves children’s ability to process information. I intend all terms within this definition to be interpreted inclusively. Mental processes include perceptual and linguistic processes, as well as conceptual, reasoning, and problem-solving ones. The improvements in children’s ability to process information that are of interest include large and small ones, long-term and short-term ones, qualitative and quantitative ones. Neural, associative, and higher-level change mechanisms are all included, because all interact to produce cognitive development.

I consider five types of mechanisms: neural mechanisms, associative competition, encoding, analogy, and strategy choice. I chose these because of their documented importance in producing cognitive growth in a wide variety of domains and over a wide age range, because they include a variety of types and levels of change, and because recent progress toward understanding them has been especially rapid. The particular labels used should not obscure the fact that many of them are best thought of as families of related mechanisms, rather than as a single one. For example, encoding is a member of a family that includes such relatives as differentiation, feature discrimination, perceptual learning, representation, and assimilation. Analogy is clearly related to induction, transfer, metaphor and simile, schema formation, and generalization. As a greater number of well-worked-out models of mechanisms become available, there will be a better basis for determining which within-family distinctions make a difference.

Perhaps the basic issue raised by the review is whether these mechanisms are best thought of as five distinct entities or as five facets of a single underlying mechanism. Among the arguments for viewing the five as separate mechanisms are that they operate in different contexts, vary greatly in grain of detail and generality of scope, and accomplish distinct cognitive functions. The main argument for viewing them as separate manifestations of a single change mechanism is that all share a core similarity. As described in greater detail in the discussion section, this core similarity involves each mechanism appearing to operate through competition among diverse cognitive entities. Here I do not resolve whether the five are best viewed as singular or as distinct
mechanisms. However, the fact that we can even raise the question and have a reasonable data base from which to discuss it attests to the value of thinking hard about mechanisms of development.¹

NEURAL MECHANISMS

Recent advances in neuroscience have provided much stronger evidence than previously existed for contributions for specific aspects of brain maturation to specific cognitive acquisitions (Cynic & Pennington 1987). Several types of neural change appear to exert especially large influences.

Synaptogenesis

The number of synapses within numerous parts of the brain follows a distinctive developmental course, in which there is initial overproduction and later pruning of synaptic connections. Such connections are produced in especially great numbers during the late prenatal and early postnatal periods. For example, Huttenlocher (1979) found that the average number of synaptic connections in the third layer of the middle frontal gyrus grew from 10,000 to 100,000 between birth and 12 months of age. The density of synapses increased until age 2, after which it gradually decreased to adult levels. These adult levels were reached by about age 7; from age 6 months to age 7 years, synaptic density in the children's brains exceeded adult levels.

Recent studies have linked cognitive and synaptic change. Turner & Greenough (1985) reported that rats housed in large complexes of cages that the animals were free to explore and that were filled with diverse objects formed 20–25% more synapses per neuron in the upper visual cortex than animals raised in impoverished environments. Chang & Greenough (1982) severed the corpus callosa of rats and then provided them with monocular experience running a maze. More extensive dendritic fields were found within a month in the occipital cortex of the hemisphere that received the input relative to the one that did not.

Advances in understanding of synaptogenesis have led to new perspectives on classic problems in cognitive development. For example, Goldman-Rakic (1987) suggested that early synaptic overproduction is critical to the development of ability to solve A-not-B object-permanence and delayed-response problems. With both monkey and human infants, ability to perform these tasks follows soon after the age at which synaptic density first exceeds adult levels, roughly age 2 months in monkeys and 6 months in humans. This

¹The strict page constraints of this volume precluded examination of several other important mechanisms. Among these are processing efficiency (e.g. Case 1985; Kail 1986), conceptual constraints (Gelman 1988; Pinker 1987), and social scaffolding (e.g. Palincsar & Brown 1984; Rogoff 1986). Future, more complete discussions of cognitive-developmental mechanisms should definitely include consideration of these ideas.
synaptic overproduction affects the prefrontal cortex and a number of areas to which the prefrontal cortex is strongly connected and that may be critical for performing such visuo-spatial memory tasks (Rakic et al. 1986). Further, surgically produced lesions in the prefrontal cortex resulted in adult rhesus monkeys' showing the same pattern of performance as infant humans and monkeys: consistent success when the object was hidden in the previously rewarded location, and near-chance performance when the location was changed (Diamond 1985). Similar-size lesions in the parietal lobe did not produce such interference on these tasks.

Goldman-Rakic (1987) theorized that the high density of synapses in the prefrontal cortex and connected areas is needed for infant monkeys and humans to initially form enduring representations capable of overcoming associative habits. She noted that the lesioned monkeys did not show perseverative errors with delay periods below 2 sec. However, at all delays equal to or greater than 2 sec, when working-memory representations would presumably be necessary for success, their performance was seriously disrupted. Performance with 2-sec delays was disrupted almost as much as performance with 10-sec delays. Overall, the monkeys with prefrontal lesions appeared unable to use working-memory representations to regulate their performance so as to transcend the effects of the previous reinforcement pattern.

Goldman-Rakic suggested that ability to form enduring representations is also crucial for other types of competence that develop substantially in the period between 6 and 18 months. In particular, she hypothesized that ability to form, access, and use on-line such representations is critical to the marked progress that children show in walking and speaking during this period. Brody's (1981) explanation of changes in short-term cued recall between 8 and 16 months; Fox et al's (1979) explanation of object permanence, stranger distress, and separation anxiety; and Fischer's (1987) explanation of complex means-ends variation in actions, use of single words, and general stage transitions in infancy have also emphasized the types of changes in brain function in general and synaptogenesis in particular that occur in this period of infancy.

**Segregation of Neuronal Input**

Synaptogenesis is not the only type of neural change that influences early development. Held and his colleagues have linked segregation of neural input from the two eyes to the development of stereopsis. Stereopsis is the perception of depth based on disparity of images impinging on each eye. The onset of stereopsis within individual infants is sudden, often developing from no detectable presence to high levels within a few weeks (Birch et al. 1982).

Held (1985) suggested that the mechanism that leads to stereopsis involves segregation of neural pathways, so that axons from the two eyes do not synapse on the same cortical neuron. Prior to the segregation of neurons into ocular dominance columns in Layer IV of the visual cortex, axons from the
two eyes often synapse on the same cells (LeVay et al 1980). When this occurs, there is no obvious way for the visual system to distinguish which information came from which eye. Held hypothesized that absence of discriminating information inhibits formation of cells sensitive to binocular disparity, since such cells would lack the information on which to operate.

Based on this view of the neural mechanism leading to stereopsis, Held and his colleagues developed an intriguing idea concerning infants’ perceptual experience. They hypothesized that prestereoptic infants nonselectively combine whatever pattern of stimulation comes into the two eyes, essentially overlaying one eye’s input on the other’s. To test this prediction, Shimojo et al (1986) longitudinally followed infants’ performance from 3 to 34 weeks on a preferential looking task. The infants wore polarizing goggles, which allowed them to receive different inputs to the two eyes. The stimulus on one side projected identical vertical lines to both eyes. The stimulus on the other side projected vertical lines to one eye and horizontal lines to the other. Due to binocular rivalry, adults see this latter combination as a constantly shifting and rather unpleasant set of line segments without intersections. Stereoptic infants were expected to experience the display similarly. On the other hand, if, as hypothesized, prestereoptic infants essentially overlayed the vertical lines presented to one eye and the horizontal lines presented to the other, they would see a checkerboard grid pattern. Infants are known to prefer checkerboard grids to vertical lines. Therefore, if prestereoptic infants experienced the vertical and horizontal lines as a grid, they would look more at them than at the vertical lines. Infants who possessed stereopsis, in contrast, would prefer the stable vertical lines to the constantly shifting line segments they would presumably experience when presented the horizontal and vertical lines.

This theoretical prediction was dramatically confirmed. When the infants were below 13 weeks, almost all of them (25 of 27) preferred the side where one eye received horizontal and the other side vertical lines. By the end of the experiment, every infant significantly preferred the side where both eyes received vertical lines. The mean age of onset for the significant preference for the vertical lines was 14 weeks, the mean age at which stereopsis also emerges. The change in preference was dramatic within individual infants. Many infants changed within 1 or 2 weeks from a significant preference for one eye receiving horizontal and the other receiving vertical lines to a significant preference for both eyes receiving the vertical lines. It seems unlikely that this difference in the visual experience of prestereoptic and stereoptic infants would have been discovered, much less predicted, without the hypothesis concerning the type of neural change that produces stereopsis.

*Experience-Expectant and Experience-Dependent Processes*

Hypotheses about change mechanisms can also lead to new perspectives on quite general issues about cognitive development. Based on their analysis of
synaptic change, Greenough et al (1987) distinguished between experience-expectant processes and experience-dependent processes. Their analysis has interesting implications both for the frequently made distinction between development and learning (Liben 1987) and for analyses of sensitive periods (Hinde 1983).

Experience-expectant processes illustrate the development end of the development-learning continuum. These processes are hypothesized to be based on the synaptic overproduction and pruning described above. Greenough et al suggested that the initial overproduction of synapses is maturationally regulated, but that which ones are pruned depends on experience. Normal experience at the normal time results in neural activity that maintains typical connections; abnormal experience, or lack of experience at the usual time, results in atypical connections. The type of experience relevant to such experience-expectant processes is experience that has been widely available throughout the evolutionary history of the species. Such experience-expectant processes tend to develop early in life and to be relatively invariant across individuals.

Greenough et al suggested that one advantage of such processes is that they allow both efficient acquisition in normal environments and reasonable adaptation to abnormal environments. In particular, the genes provide a rough outline of the eventual form of the process, thus allowing quite rapid acquisition under usual circumstances. Unusual environments or physical deficiencies, however, lead to unusual neural activity, which creates alternative neural organizations that are adaptive given the unusual circumstances.

The usefulness of this concept of experience-expectant processes for understanding sensitive periods can be illustrated in the context of the segregation of neural input from each eye, described above. In young cats and monkeys, sewing shut one eye during an early sensitive period results in a severe visual impairment when the eye is later reopened (Wiesel & Hubel 1965). Instead of the usual process, in which terminal fields of axons from both eyes are cut back equally from the visual cortex columns dominated by the other eye, the terminal fields of the deprived eye are cut back from a larger area, and those from the nondeprived eye from a smaller area, than usual. The abnormal experience leads to neurons from the nondeprived eye winning competitions that these neurons are predisposed to lose when both eyes receive typical input.

The timing of the sensitive period seems to be a function both of when synaptic overproduction occurs and when the organism receives relevant experience. In kittens, the greatest damage from monocular deprivation occurs when the deprivation is in the first two months of life, overlapping substantially with the beginning of the period of synaptic overproduction. Conversely, even small amounts of binocular experience in the first two months are sufficient to protect the kitten from harmful effects of later
monocular deprivation (Hubel & Wiesel 1970; Mower et al. 1983). However, dark-reared kittens who do not receive any early visual experience are still harmed by monocular deprivation when it is introduced at 10 months (Cynader & Mitchell 1980).

To explain both the age-relatedness of the sensitive period and the lack of rigidity of its age boundaries, Greenough et al. suggested that visual experience has the effect of committing a set of synapses to a particular organization. Once this commitment has occurred, synapses not needed for that organization are pruned away. If no relevant experience is encountered, the synaptic pruning is forestalled for a number of months. As Bertenthal & Campos (1987) noted, this analysis of sensitive periods advances understanding beyond the usual criteria of age-dependency and irreversibility. It indicates what produces the sensitive periods, why they usually occur when they do, and how, under certain circumstances, they can be prolonged.

The other side of Greenough et al.'s dichotomy involves experience-dependent processes. These are the neural substrate of what is usually thought of as learning. With experience-dependent processes, formation of synaptic connections depends on experiences that vary widely among individuals in whether and when they occur. The experience-dependent processes appear to operate through the formation of synapses in response to specific neural activity caused by partially or totally unsuccessful attempts to process information. Such synapses can be generated as rapidly as 10–15 min after a new experience (Chang & Greenough 1984). Synapse production under such circumstances appears to be fairly localized to the site of the previous information processing. However, as with experience-expectant processes, more synapses than will later be present may be produced. The synapses that are maintained are those involved in subsequent neural activity.

These analyses suggest both similarities and differences between experience-expectant and experience-dependent processes. In both, the change mechanism involves a cycle of synaptic increase and pruning. Also in both, neural activity determines which synapses are maintained. However, the events that trigger the synaptogenesis, its degree of localization, the ages at which it occurs, and the behavioral domains influenced all distinguish the two types of process. Overall, the emphasis on neural change mechanisms allows a considerably more differentiated approach to the relation between learning and development than has previously been possible.

ASSOCIATIVE COMPETITION

Synaptic connections that are used are maintained; those that are not used are pruned. This phenomenon has provided an analogy from which a variety of hypotheses about associative competition have been drawn. The basic idea of
the outcomes of cognitive competitions leading to changes in children's associative networks, and therefore in their thinking, has been hypothesized to contribute to language development (Bates & MacWhinney 1987; MacWhinney 1987; Rumelhart & McClelland 1986), changes in the organization of free recall (Bjorklund & Jacobs 1985), acquisition of arithmetic and reading skills (Ashcraft 1987; Seidenberg & McClelland 1988), and improvements in problem solving (Anderson 1987; McClelland 1989). Some of the best-worked-out ideas about how associative competition might operate have been developed in the context of connectionist models. Therefore, these are emphasized in the present discussion.

Connectionist Models

Connectionist models, emphasizing distributed representations, parallel processing, and interactions among large numbers of simple processing units, have greatly influenced recent thinking about many aspects of adult cognition (Hinton & Anderson 1981; McClelland & Rumelhart 1986; Palmer 1987). Not surprisingly, such models are beginning to influence ideas about cognitive development as well.

Most current connectionist models involve an input level, one or more hidden levels, and an output level. Each of these levels contains a number of discrete units. For example, in MacWhinney and colleagues' (Taraban et al. 1989) model of how German children acquire the set of articles used in their native language ("der", "die", "das", etc), the input level includes 39 units, the output level contains 6 units, and the two hidden levels have 30 and 7 units, respectively. The 39 input units represent individual auditory, lexical, case, and semantic features of the input noun. The 2 levels of hidden units represent a variety of combinations of these input-level features. The 6 output units represent the 6 articles in the German language. Each input-level unit is connected to each of the units in the hidden level immediately above it, each of these units is connected to each of the units at the next higher hidden level, and each unit at that level is connected to each output unit.

The processing that goes on within connectionist models can also be described in the context of the Taraban et al simulation. The simulation represents the noun that is presented in terms of a subset of the 39 input-level units. Those input-level units that are active feed activation into the units of the hidden level immediately above it. The activation that a given hidden-level unit receives from a given input-level unit is a function of whether the input-level unit is active and of the strength of the connection between input and hidden unit. The total activation received by a given hidden unit is the sum of the positive and negative activations contributed to it by all of the input units that are connected to it. The activation of each output-level unit is produced in the same way, except that the immediate source of the activation is hidden-level rather than input-level units.
Now consider how learning occurs in such models. The particular learning algorithm used by Taraban et al and in most other recent connectionist models involves backward propagation of error corrections. Such “back propagation” begins by the activations of the output units being compared to those of a target that is intended to represent feedback from the environment. For example, on a given trial, the simulation’s processing might result in “der” receiving an activation of .4, and “die” an activation of .5, when “der”, the correct article for that noun, would ideally have received an activation of 1, and “die,” an incorrect article, an activation of 0. For each output-level unit, the value of the desired and actual activations would be compared to obtain a measure of the degree of error. The strengths of connections from all units that sent activation to higher level units would then be adjusted so that on subsequent presentations of the stimulus they would produce values of the output unit that more closely approximated the desired state. For example, strong connections between a hidden-level unit and an incorrect output-level unit would be decremented, whereas weak connections between a hidden-level unit and the correct output-level unit would be strengthened.

**Associative Competition and Language Development**

What can such associative competitions accomplish? Taraban et al chose to test the ability of the MacWhinney (1987) competition model to acquire the article system of German. This task was of interest precisely because the German article system is so difficult. The appropriate article varies with several properties of the noun: gender (masculine, feminine, or neuter), case (nominative, genitive, dative, or accusative), and number (singular or plural). To make matters worse, assignment of nouns to gender categories is often nonintuitive: For example, the word for “fork” is feminine, the word for “spoon” is masculine, and the word for “knife” is neuter. Maratsos (1982) argued that neither semantic nor phonological cues predict which article accompanies a given noun, and that only purely syntactic cues could allow correct choices to be made.

In contrast, Taraban et al sought to demonstrate that available lexical, semantic, phonological, and case cues were sufficient for children to learn the German article system. Input to the system was 102 common German nouns, each presented a number of times. The system needed to choose which article to use with that noun in the particular case-and-number context.

After experience with this training set, the Taraban et al simulation chose the correct article for 98% of the nouns. This could not be attributed simply to rote learning of the possible article-noun combinations. When the simulation was presented a previously encountered noun in a novel case role, it chose the correct article on 89% of trials. The simulation also proved able to generalize to entirely novel nouns. The 48 most frequent nouns in German that had not been included in the original input set were presented in all possible case
roles. On this completely novel set, the simulation chose the correct article from among the 6 possibilities on 63% of trials, versus 17% expected by chance. Thus, the system's learning mechanism allowed it to make a good guess about what article would accompany a given noun, even when the noun was entirely unfamiliar.

The simulation's learning resembled children's in a number of ways. Early in acquisition, it, like children, tended to overuse the article that accompanies feminine nouns. Similarly, both the simulation and children tended to acquire the connection between the -e noun ending and feminine articles early in learning. Further, the same article-noun combinations that are the most difficult for children proved the most difficult for the simulation.

MacWhinney (1989; Bates & MacWhinney 1987) proposed that two key determiners of the difficulty of acquisition in such a system are cue availability (the frequency with which a cue is present) and cue reliability (the predictive accuracy of the cue when it is present). Results obtained from the Taraban et al simulation indicated that cue availability is a good predictor of relative speed of learning early in the learning process, but that cue reliability becomes the best predictor later in learning. Very late in learning, conflict-validity, the reliability of a cue when other cues point to different answers, may become an even better predictor of difficulty than general cue reliability (McDonald & MacWhinney 1989). The pattern suggests that as learners obtain increasing amounts of input, they rely increasingly on specifically relevant information, at the expense of less-relevant but more easily available information. This seems a plausible sequence for many aspects of cognitive growth, both in language and in other areas.

ENCODING

Although connectionist accounts are very promising, existing connectionist models also have certain limitations as explanations of development (Pinker & Prince 1988). One main limitation is that they assume veridical encoding of stimuli at the input level from the beginning of learning. For example, as described above, the Taraban et al simulation assumes that from trial 1, nouns are encoded on the 39 input features involving gender, case roles, phonology, and meaning. In human development, however, children often fail to encode relevant features; much of cognitive growth involves the acquisition of increasingly adequate encodings. Comprehensive accounts of development must explain these changes in encoding, as well as changes in how encoded information is used.

Encoding is the process by which stimuli are represented in a particular situation. In almost any context, some information is encoded and used to operate on the environment; other information is encoded but not used; yet other information is not encoded. When existing strategies fail, information
that is encoded but not used provides a kind of reserve capital for constructing new, potentially superior approaches. For example, a boy approaching a playground on his first day at school may encode the presence of a friend and use this information in deciding to go over to play with the friend; may encode the presence of a playground monitor in another part of the playground but make no use of the information; and may not encode the fact that all of the older children are playing near his friend and all of the other young children are playing in a different part of the playground. If the older children started bothering the boy, he could use his encoding of the playground monitor’s location to go over to her for safety, and could learn to stay near her in the future. However, not until he encoded that the older and younger children generally played in different parts of the playground could he learn to avoid playing where the older children were, and thus free himself of the need to monitor the monitor. Thus, encoding potentially relevant information can facilitate learning, and failing to encode relevant information can hinder or preclude learning.

Recent Empirical Findings

Changes in encoding have been found to underlie a variety of age-related changes in children’s thinking. When presented balance scale problems, 5-year-olds encode only the weight dimension whereas 8-year-olds encode both weight and distance; teaching the 5-year-olds to encode distance as well as weight allows them to learn new rules from experiences that previously were not helpful (Siegler 1976). Adults spend more time than 7-year-olds encoding analogical reasoning problems; this lengthier encoding is more than compensated for, however, by the much more rapid execution of other components that the more extensive initial encoding makes possible (Sternberg & Rifkin 1979). Older children encode transitive inference problems in relative terms (“A is greater than B”) whereas 6- and 7-year-olds encode them in absolute terms (“A is long and B is short”); the younger children may have difficulty solving transitive inference problems because their absolute encodings lead to inconsistencies, such as B being encoded as long on one trial but short on another (Perner & Mansbridge, 1983; Riley & Trabasso 1974). Across a wide range of tasks, older children’s encoding has been found to be both more selective and more exhaustive of the relevant features than that of younger children (Sophian 1984; Sternberg & Powell 1983).

Differences in encoding are also related to individual differences in the thinking of children of a single age. Differences between high-IQ and average-IQ 7-year-olds’ analogical reasoning parallel those between adults and 7-year-olds. The higher-IQ children take more time to encode but then solve the problem more quickly (Sternberg & Rifkin 1979). Gifted 9- to 11-year-olds are more likely than age peers to focus their encoding on the critical parts of insight problems and to ignore irrelevant parts (Davidson 1986; Marr &
Sternberg 1986). Independent of IQ, musically gifted 12- and 13-year-olds encode verbal material more efficiently than do artistically gifted peers, whereas the artistically gifted children encode visual material more efficiently (O’Connor & Hermelin 1983). Word encoding of 11-year-olds with learning disabilities, like that of typical 6-year-olds, emphasizes acoustic features, whereas word encoding of typical 11-year-olds emphasizes semantic features (Lorsbach & Gray 1985). The importance of encoding also extends to social domains; for example, aggressive 7- to 9-year-olds encode other children’s intentions less accurately than do less aggressive peers (Dodge et al 1986).

How New Encodings Are Formed

Holland and his colleagues (Holland 1986; Holland et al 1986) suggested a genetic algorithm by which new encodings, and hence new opportunities to learn, might be generated. Within Holland’s model, a learner’s knowledge is represented as a set of rules. Each rule is a condition-action pairing, much like a production in a production system. Within a rule, a string of 1s, 0s, and neutral values on the condition side indicates the rule’s encoding of the environment. Another string of 1s, 0s, and neutral values on the action side indicates what changes occur when that rule fires. Each rule also has a strength, which determines its probability of firing when several rules are applicable. Rule strengths change as a function of how often the rule is used and how consistently its use is followed by attainment of the system’s goal.

Within Holland’s genetic algorithm, new encodings and rules are produced when, periodically, two rules with different condition sides but identical action sides are selected to be “parents.” For example, two parent rules might be: 101→010 and 000→010. New rules are produced by making a cut at an arbitrary point in the condition side of each parent, combining the part of the condition to the left of one cut with the part of the condition to the right of the other cut, and linking that new condition side to the action side of the parent productions. In the above example, making a cut between the second and third symbol on the condition side of each parent rule would give rise to the new rules: 100→010 and 001→010.

The probability of a given rule’s being chosen to be a parent is proportional to its strength. This assures that rules more closely linked with goal attainment more often contribute to new encodings. It also assures that other rules have some chance of doing so, which provides potentially useful variability in the rules that are formed. Whenever new rules are created, they replace the weakest of the existing rules. The set of rules is thus driven by increasingly useful encodings.

Although this approach does not seem to have been applied yet to modeling children’s thinking, it has illustrated how improved encodings could contribute to cognitive change in a number of interesting contexts, among them poker playing, Prisoner’s Dilemma games, and gas-pipeline transmission
problems (Holland et al 1986). Its learning to solve the pipeline problems is particularly impressive, because the system’s initial encodings and rules were generated randomly and thus did not assume any prior knowledge of the environment. Input to this simulation was hourly information from a simulated gas transmission system concerning inflow and outflow of gas, input and output pressure of gas, rate and direction of changes in pressure, date, time, and current temperature. The system’s task was to deliver gas at the minimum pressure needed to meet demand and to locate any leaks that were present. The genetic algorithm formed new encodings and rules after every 200 hours of experience.

Despite the initial random generation of rules (and encodings), the system progressed to near optimal performance in detecting gasoline leaks and compensating for them by sending additional gas through that part of the pipeline. It eliminated unreasonable rules, restricted the range of overly general rules, and produced new rules that were more useful than any of the previous ones. For example, it eventually produced the rule “If input and output pressure are low and the rate of change in pressure is very negative, then send a ‘leak’ message.” This rule gained considerable strength once it was generated, because the simultaneous presence of the three encoded states was very predictive of actual leaks. The simulation demonstrates how a “mindless” mechanism for generating new encodings can come to encode the environment in new and useful ways.

ANALOGY

Analogy produces progress in children’s thinking by allowing them to interpret poorly understood situations in terms of better-understood ones. It can be especially helpful in problem solving, where a novel (target) problem is often understood in terms of a familiar (base) problem. For example, people are more likely to solve Duncker’s famous X-ray problem (focusing separate X-rays at the location of a tumor) if they have already learned the solution to a parallel problem in which an attacking army must divide into separate attack units and travel from different directions to converge at the enemy’s location (Gick & Holyoak 1980).

Recent Empirical Findings

Knowledge about the development of analogical reasoning has expanded rapidly in the past few years. Children as young as 3–5 years have been found to solve new problems more effectively if they had previously solved analogous problems (Brown et al 1986; Chen & Daehler 1988; Crisafi & Brown 1986; Holyoak et al 1984). This is true even when there is little perceptual similarity between the base and target problems (Brown et al 1986; Crisafi & Brown 1986). Having the experimenter and the child state the common solution principle that unites the two problems (Crisafi & Brown 1986),
presenting problems in an order where the causal structures proceed from most to least obvious (Crisafi & Brown 1986), and prompting children to recall key elements of the original problem’s causal structure (Brown et al 1986) all help 3- to 5-year-olds to analogize from familiar to novel problems. Even 2-year-olds show analogical transfer when they understand the causal relations in both situations (Brown 1989).

Whether children detect potentially useful analogies depends on a variety of characteristics of the problem situation and of the child’s cognitive activities. Among the potentially helpful aspects of the problem situation are similar base and target problems (Brown & Campione 1984; Gentner & Landers 1985; Holyoak et al 1984), presentation of multiple base problems with the same solution principle (Crisafi & Brown 1986; Gholson et al 1987), and provision of visual representations diagramming the essential aspects of the problem (Beveridge & Parkins 1987). Among the key aspects of children’s cognitive activities are completeness of encoding (Chen & Daehler 1988; Sternberg & Nigro 1980; Sternberg & Rifkin 1979) and amount of knowledge of the base problem (Inagaki & Hatano 1987). These features seem to be influential at all ages.

Young children’s ability to draw analogies, and their sensitivity to many of the same problem characteristics that influence the analogical reasoning of older children and adults, should not obscure the profound developmental changes that occur in analogical reasoning. Young children require explicit hints to draw analogies that older children draw without such hints (Crisafi & Brown 1986; Brown et al 1986). They at times draw unsound analogies to just-presented problems at alarmingly high rates (Chen & Daehler 1988). Their analogizing is hindered by superficial perceptual dissimilarities that do not influence the analogizing of older children and adults (Gentner & Toupin 1986; Goldman et al 1982; Holyoak et al 1984). Their execution of processes involved in analogizing, such as encoding, mapping, and inference, is less efficient than that of older children (Bisanz 1979; Pellegrino & Goldman 1983; Sternberg & Rifkin 1979). Thus, satisfactory models of analogical reasoning must account for a variety of age-related differences, as well as a variety of similarities.

**How Analogies Are Generated**

Gentner’s structure-mapping engine (Falkenhainer et al 1986; Gentner 1989) illustrates how well-worked-out models of performance can yield interesting insights about development. The model’s focus is on how children establish correspondences between a base domain and a target. The central idea is that children, like adults, strive to form analogies where the system of relations in the target domain resembles the system of relations in the base. The objects being related need not have any particular resemblance. Instead, the key is the similarity of the corresponding relations in the two situations.
Gentner (1989) used Carnot’s analogy “heat flow is like water flow” to illustrate the workings of the structure-mapping model. The concrete situation she used to illustrate heat flow involved a cup of hot coffee, an ice cube, and a silver bar connecting the two. The situation used to illustrate water flow involved a large beaker with a tall liquid column, a small vial with a short liquid column, and a pipe running between the bottom of the beaker and the bottom of the vial. The task was to map the relations involved in how temperature differences influence the flow of heat between coffee and ice cube onto the relations involved in how pressure differences influence the flow of water between the two containers. Thus, the coffee cup needed to be mapped onto the large beaker, the ice cube onto the vial, the silver bar onto the pipe, and heat onto water.

Falkenhainer et al (1986) implemented a computer simulation to illustrate how the theory would draw this, and a number of other, analogies. Figure 1 illustrates the simulation’s initial representation of the base and target situations in the “Heat flow is like water flow” analogy. The simulation first tries to establish local matches, then assigns evidence scores to each of the local matches, then constructs global matches, and finally evaluates the global matches.

To construct local matches, the simulation examines each object and relation in the base (better-known situation) and identifies a set of objects and relations in the target (less-well-known situation) that it could plausibly match. For example, if relations in the base and target have identical names, then a match hypothesis is created. Each match hypothesis leads to checking corresponding arguments of the relations; if they are of the same type (e.g. both are functions), then a match hypothesis between them is created. This

![Diagram](image)

**Figure 1** Structure-mapping engine's representation of “heat flow is like water flow” analogy (from Gentner 1989).
often leads to a large number of local matches, some of which prove useful and others not. For example, in Figure 1, the relation GREATER places into correspondence “pressure” in the water flow situation and “temperature” in the heat flow situation (a relation that is part of a high-quality analogy) and also “diameter” in the water flow and “temperature” in the heat flow situations (a relation that is not).

Rules assign evidence scores to these local matches. One such rule is to increase the score if corresponding relations in the base and target have the same name. Another is to increase the score for a given local match if the relation immediately higher in each hierarchy also matches. This second rule leads to the preference for the GREATER relation involving pressure over that involving diameter. The higher-level “CAUSES” relation (Figure 1) adds to the evaluation of the pressure-temperature relation but not to that of the diameter-temperature relation.

To construct global matches, the structure-mapping engine combines match hypotheses into the largest possible systems whose object mappings are internally consistent. These systems are possible interpretations of the analogy. Each system also includes a set of possible inferences about how additional, unmatched parts of the base could be mapped onto the target. For example, the CAUSE relation between pressure and water flow in the base analogy leads to an inference that there may be a CAUSE relation between temperature and heat flow in the target analogy. The global matches are evaluated on the basis of the sum of the evidence from the local matches.

This structure-mapping model has proved useful for understanding how analogical reasoning develops. Falkenhainer et al described three variants of the structure-mapping engine: one that operated solely on object matches, one that operated solely on relational matches, and one that used both. These alternative versions seemed to correspond to developmental differences in analogical reasoning. Gentner & Toupin (1986) found that 4- to 6-year-olds saw situations as analogous only when corresponding objects were similar. Their behavior was like that of the object-matching simulation. In contrast, 8- to 10-year-olds did not require such similarities between objects to draw the analogies; parallel sets of relations were sufficient for them to do so. Thus, their performance was more like that of the programs that relied on relations.

Gentner (1988) found a similar developmental trend in interpretation of metaphors. Children could correctly interpret metaphors based on similarity of objects before they could interpret metaphors where only relational structures were parallel. Further, with age, children became increasingly likely to interpret relationally those metaphors that could be viewed either in terms of similarities between objects or similarities between relations. In sum, the structure-mapping engine illustrates how a variety of characteristics of analogical reasoning and its development could arise. [See Bakker & Halford (1988) for a related model of the development of analogical reasoning; this
model has the additional virtue of accounting for the age-related trend toward
detecting increasingly less obvious analogies.]

**STRATEGY CHOICE**

People can approach almost any task in multiple ways. For example, the
majority of first graders have been found to use at least three strategies in
solving a set of simple addition problems (Siegler 1987). Typically, children
counted from one on some problems, counted from the larger addend on other
problems, and retrieved an answer from memory on yet other problems. For
children, such use of multiple strategies has great advantages; it allows them
to fit their strategies to constantly changing knowledge and situational de-
mands. For researchers, however, it raises a great many questions: in particu-
lar, when do children use various strategies, how do children decide to use
one strategy rather than another, and what developments lead to changes in
strategy use with age and experience?

**Recent Empirical Findings**

From quite young ages, children choose among strategies in ways that yield
desirable combinations of speed and accuracy. This adaptive quality can be
seen in choices between retrieval and backup strategies (approaches other than
retrieval, such as looking up words in a dictionary in spelling, counting from
the hour by 5s in time-telling, repeatedly adding in multiplication, and
sounding out words in reading). Both retrieval and backup strategies have
clear, though different, advantages. Retrieval can be executed more quickly;
backup strategies can yield correct answers on problems where retrieval
would be inaccurate.

Siegler & Shrager (1984) found that in simple addition, even preschoolers
chose between stating a retrieved answer and using a backup strategy in a very
reasonable way. On easy problems, the 4- and 5-year-olds relied primarily on
retrieval; on more difficult problems, they relied primarily on backup stra-
egies. This allowed them to answer the easier items quickly and accurately,
and to answer the more difficult items accurately, though not so quickly.
Similarly strong relations between problem difficulty and strategy choices
have been found for 4- to 12-year-olds in subtraction, multiplication, word
identification, time-telling, and spelling (Geary & Burlingham-Dubree 1989;

The fact that individual children of a wide range of ages use multiple
strategies and choose among the strategies in adaptive ways does not mean
that the particular strategy choices that children make remain constant. Chi-
dren continually learn new strategies and change their frequency of use of
existing strategies (Goldman et al 1989). Age-related differences are also
apparent in children's transfer of recently learned strategies to new situations. Adults are more likely than fifth and sixth graders to transfer strategies, especially under conditions where no specific information about the relevance of the strategy is given (Pressley et al. 1984a). This is true even when both children and adults explicitly recognize that the new strategy is more effective (Pressley et al. 1984b). Older children are also more likely than younger ones to try additional strategies when their initial strategy does not produce entirely correct performance (Ceci & Howe 1978).

The Siegler (1988b) strategy-choice model illustrates how a single mechanism can give rise to both the adaptive quality of children's strategy choices at any one time and the changes in their speed, accuracy, and strategy choices that take place over time. The simulation can be divided into a representation of knowledge and a process that operates on that representation to produce performance and learning.

The representation includes knowledge of problems, of strategies, and of the interaction between problems and strategies. Knowledge of problems is represented as associations between each problem and possible answers to that problem, both correct and incorrect. For example, 5+3 would be associated not only with 8 but also with 6, 7, and 9. These representations of knowledge of each problem can be classified along a dimension of the peakedness of the distribution of associations. In a peaked distribution, most strength is concentrated in the correct answer. At the other extreme, in a flat distribution, strength is dispersed among several answers, with none of them forming a strong peak. For example, in Figure 2, the strengths for answers to 2+1 form a peaked distribution (with the strength for 3 at the peak) and those for 3+5 form a flat distribution.

The simulation's representation also includes knowledge about strategies. Each time a strategy is used, the simulation gains information about that strategy's speed and accuracy. This information generates a strength for each strategy, both in general and on particular problems. The strategies modeled in the current version of the addition simulation are the three most common approaches that young children use to add: counting from one, counting from the larger number, and retrieval.

One further feature of the new simulation's representation should be mentioned. Newly generated strategies possess "novelty points" that temporarily add to their strength and thus allow them to be tried even when they have little or no track record. The strength conferred by these novelty points is gradually lost as experience with the strategy provides an increasingly informative database about it. This feature was motivated by the view that people are often interested in exercising newly developed cognitive capabilities (Piaget 1951), and by the realization that without a track record, a newly developed strategy would be unlikely to be chosen if previously acquired strategies worked reasonably well.
Figure 2  A peaked (left) and a flat (right) distribution of associations. The values for each answer reflect the percentage of children who stated that answer in an experiment in which children needed to state the first answer to the problem that they thought of (the retrieval-required experiment reported in Siegler 1986).

Now consider the process that operates on this representation to produce performance. First, a strategy is chosen. The probability of a given strategy being chosen is proportional to its strength relative to the strength of all strategies. (Recall that this strength reflects the speed and accuracy the strategy has generated in previous uses.) If a strategy other than retrieval is chosen, that strategy is executed. If retrieval is chosen, the simulation retrieves a specific answer (e.g. 4). Just as the probability of a given strategy being chosen is proportional to its strength relative to the strength of all strategies, the probability of any given answer being retrieved is proportional to its strength relative to the strength of all answers to the problem. Thus, in Figure 2, the connection between “2+1” and “3” has a strength of .80, the strength of connections between “2+1” and all answers is 1.00, so the probability of retrieving “3” is 80%. If the strength of whichever answer is chosen exceeds the confidence criterion (a threshold for stating a retrieved answer), the simulation states that answer. Otherwise, the simulation again cycles through the strategy-choice process; it does so until a strategy is chosen and an answer stated.

At any one time, the simulation generates patterns of accuracy, solution times, and strategy use much like those of children. For example, it counts on from the larger number most often on problems where the smaller of the two addends is small and where the difference between the two addends is large. Siegler (1987) found the same pattern in kindergarteners’, first graders’, and second graders’ performance. Also as with children, the simulation uses
retrieval most often on problems where both addends are small and uses counting from one primarily on problems where both addends are large. Relative problem difficulty and particular errors that the simulation makes also parallel those of children.

*How Strategy Choices Change*

The simulation learns a great deal through its experience with strategies and problems. As it gains experience, it produces faster and more accurate performance, more frequent use of retrieval, less frequent use of counting from one, and closer fitting of when strategies are used to their advantages and disadvantages on each problem. For example, the simulation’s accuracy improves from 55% correct relatively early in its run to 99% correct by the end. Similarly, it progresses from using retrieval on 16% of trials at the earlier point to 99% at the end. Such learning is produced entirely through children associating strategies with the speed and accuracy that they have produced and associating answers with the problems on which those answers have been stated.

The way that this learning mechanism operates can be illustrated in the context of how some strategies come to be chosen on some problems more often than others. Consider two problems, 9+1 and 5+5. Kindergarteners and first graders count on from the larger number considerably more often on 9+1, yet count from one more often on 5+5. The simulation generates similar choices between the strategies and illustrates how such a pattern might emerge. On 9+1, counting from the larger number has a very large advantage in both speed and accuracy over counting from one. It requires only 1/10 as many counts. In contrast, the numbers of counts required to execute the two strategies are more comparable on 5+5, where counting on from the larger number requires 1/2 as many counts. If the number of counts were the only consideration, children might be expected to consistently count from the larger number on both problems (and on all problems) from the time they learned how to do so. However, for any given number of counts, counting on from an arbitrary number is considerably more difficult for young children (in terms of time and errors per count) than counting from one. The simulation’s probability of erring on each count and its time per count reflect this greater difficulty of counting on from a number larger than one. Thus, the simulation learns that although counting from the larger number is generally more effective, there are some problems, such as 5+5, where counting from one works better. This leads to counting from one being the most frequent strategy on such problems for a period of time.

Eventually both counting from one and counting from the larger number are overtaken by retrieval. At the outset of the simulation, all answers to each problem have similar, minimal associations with the problem; the distribution
of associations is flat. The more often that children encounter problems, and the more accurate the execution of backup strategies (strategies other than retrieval), the more the strength of the correct answer grows relatively to the strength of incorrect answers. Thus, as the simulation gains experience, the distribution becomes increasingly peaked. As discussed in detail in Siegler & Shrager (1984), the more peaked a distribution of associations, the more likely the child is to retrieve the correct answer and the more likely that answer is to exceed the confidence criterion and therefore to be stated. Thus, as children generate the correct answers to problems increasingly often, they become increasingly likely to solve the problems through retrieval.

In addition to illustrating how children in general may choose strategies, the strategy-choice model also has proved useful for analyzing individual differences. Siegler (1988c) examined first graders’ consistency of strategy choices across addition, subtraction, and word-identification tasks. Cluster analyses of the children’s performance suggested that they fell into three groups: “good students,” “not-so-good students,” and “perfectionists.”

The contrast between good and not-so-good students was evident along all of the dimensions that might be expected from the names. The good students were faster and more accurate on both retrieval and backup strategy trials on all three tasks, and also used retrieval more often on all of them.

Whereas the differences between good and not-so-good students could be ascribed to simple differences in knowledge, the differences between good students and perfectionists could not be. The two groups were equally fast and accurate on both retrieval and nonretrieval trials. However, the perfectionists used retrieval significantly less often than not only the good students but even the not-so-good students.

The differences among good students, not-so-good students, and perfectionists could be interpreted in terms of two key variables within the model: the peakedness of distributions of associations, and the stringency of confidence criteria. In particular, perfectionists appeared to be children who possessed peaked distributions and who set very high confidence criteria; good students appeared to be children who also possessed peaked distributions but who set somewhat lower confidence criteria; not-so-good students appeared to be children who possessed flat distributions and set low confidence criteria. Within the computer simulation, these combinations of distributions of associations and confidence criteria produced performance much like that of the three groups of children. For example, the perfectionists’ peaked distributions and high confidence criteria would lead to fast and accurate performance when they used retrieval, but also to relatively little use of retrieval.

Subsequent achievement test scores and class placements provided external validation for the experimental analysis (Siegler 1988c). Perfectionists and good students scored equally highly on standardized mathematics and reading
achievement tests that were given four months after the experimental sessions; perfectionists’ scores averaged at the 81st percentile, good students’ at the 80th percentile. Children in both groups scored much higher than children classified as not-so-good students, whose achievement test scores averaged at the 43rd percentile. Further, 4 of the 9 children in the not-so-good students group either were assigned the next year to a learning disabilities class or retained in the first grade, versus 0 of 27 perfectionists and good students. Thus, the differences between not-so-good students and the other two groups were apparent in conventional measures of individual differences as well as within the analyses suggested by the model. However, the differences between good students and perfectionists would probably have gone undetected without the analysis of the strategy choice mechanism.

CONCLUSIONS

These ideas and results illustrate how detailed analyses of change mechanisms can contribute to understanding of cognitive development. The findings of Held and his colleagues about the perceptual experience of pre-stereoptics infants; those of MacWhinney and his colleagues about the learnability of the German article system on the basis of phonological, case, and lexical cues; and my own findings about the strategy choices of good students, not-so-good students, and perfectionists illustrate the nonintuitive empirical findings that can arise from such analyses. Gentner’s distinction between analogies based on object and relational correspondences, Greenough’s concepts of experience-expectant and experience-dependent processes, and Holland’s genetic algorithm illustrate how children can quickly acquire initial understanding of phenomena, yet take a very long time to acquire advanced understanding of them. Many of these analyses, as well as Goldman-Rakic’s analysis of the development of ability to form enduring representations, illustrate how a single mechanism can contribute to a large variety of behavioral changes.

If we think about these five change mechanisms as a group, and ask what they have in common, a single theme keeps on recurring. This theme is competition. The competitions are sometimes among synaptic connections, other times among associative processing units, other times among problem-solving strategies. Sometimes they occur in the context of perceptual development, other times in language development, other times in the development of analogical reasoning and problem solving. Constant across the hypothesized mechanisms, however, are two essential features of competition: multiple competing entities and a means for choosing among them.

Consider the types of competitions that characterize the mechanisms reviewed in this chapter. With regard to synaptogenesis, the competition involves an early density of synapses greater than that in mature brains. Normal
experience at the normal time results in typical connections' being maintained. Abnormal experience, or lack of relevant experience, results in atypical patterns, in which synaptic connections win competitions that they ordinarily would lose. The view that neural activity determines the outcomes of synaptic competitions is evident in Greenough's experience-expectant and experience-dependent processes, and also in Held's model of the development of binocular-disparity detection cells.

Competition plays a similarly central role in connectionist models. Such models include huge numbers of connections among processing units. Back propagation provides a way of raising the strength of connections that provide useful input to units at the next higher level, and reducing the strength of connections that do not. Thus, MacWhinney (1987) had good reason to label his connectionist model of language development "The Competition Model."

Competition also plays a prominent role in ideas about how encoding might be accomplished. At any one time, a number of aspects of situations are encoded, both ones currently being used and ones that are not. Over time, encodings that prove advantageous come to be used increasingly often. For example, within Holland et al.'s genetic algorithm, numerous encodings compete to be the parents from which new encodings are formed. The likelihood that a given encoding becomes a parent is determined by the strength of the rule of which that encoding is a part. Rules (and encodings) that do not lead to attainment of goals lose competitions and are eventually deleted; those that prove useful are employed increasingly often.

Alternative analogies seem to compete much as alternative encodings do. An infinite number of analogies can be drawn, but only the ones that create systematic correspondences are highly evaluated. Consistent with this view, Gentner's structure-mapping engine detects a wide range of parallels between relations in base and target domains, but selects only those mappings that create the most systematic correspondences between relations in the two situations.

The whole idea of strategy choice is based on the notion of competition among multiple strategies. Rather than consistently using one strategy at an earlier age and a different strategy at a later one, children often use multiple strategies at both earlier and later ages. The Siegler strategy-choice model describes how the speed, accuracy, and answers previously generated by each strategy shape strategy choices at any one time and also lead to changes over time in which strategies are chosen.

Competition may also be at the core of another developmental mechanism not discussed here: Piaget's concept of equilibration. Within the equilibration construct, there is continuous competition between representations of new events and longer-term cognitive structures. At times of stage transition, preexisting and emerging cognitive structures also compete. The goodness of
fit between the inherent structure of the environment and the child’s alternative representations of the environment determines which cognitive structures prevail.

In all of these cases, competition seems to serve the same function. The multiple competing entities provide the variability needed to adapt to changing environments, contextual demands, and organismic capabilities. The selection methods produce cognitive growth by leading to greater use of those processing units that have proved useful under the particular circumstances encountered by the child. These variability and selection functions seem essential to any developing system. Thus, they may be a basic part of many, if not most, mechanisms of cognitive development.

ACKNOWLEDGMENTS

This research was supported in part by grant HD-19011 from the National Institutes of Health, in part by a grant from the Spencer Foundation, and in part by a grant from the McDonnell Foundation. Thanks are due to Robert Glaser, Patricia Goldman-Rakic, Carl Granrud, William Greenough, David Klahr, Brian MacWhinney, Jay McClelland, Mitchell Rabinowitz, and James Staszewski for comments on earlier versions of the manuscript. Request for reprints should be sent to Dr. Robert Siegler, Psychology Department, Carnegie-Mellon University, Pittsburgh, PA 15213.

Literature Cited


Case, R. 1985. Intellectual Development;
COGNITIVE DEVELOPMENT 377


Holyoak, K. J., Junn, E. N., Bilman, D. O. 1984. Development of analogical problem-
solving skill. Child Dev. 55:2042–2055
Miller, P. H. 1983. Theories of Developmental Psychology. San Francisco: Freeman
Siegler, R. S. 1988c. Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. Child Dev. 59:833–51
Siegler, R. S. 1989. How domain-general and domain-specific knowledge interact to produce strategy choices. Merrill-Palmer Q.