change: genetic drift, bottleneck and founder effects, assortative mating, and environmental selection. Finally, advances in genetics have profoundly increased understanding of the mechanisms of inheritance, as well as all of the other evolutionary functions.

Breakthroughs in understanding the immune system provide a concrete instance of how understanding of changes within an individual's lifetime can benefit from application of these evolutionary concepts. Thinking about the immune system within an evolutionary framework has facilitated recognition of the immense, spontaneously occurring variability that exists in the population of white blood cells that produce immune responses. It also has promoted discoveries regarding the self-regulation and adaptive change that occurs through exposure to an antigen giving rise to increased number of cells that react to that antigen, and of the inheritance produced by memory cells that still react to the antibody decades later. The hope is that thinking of cognitive development within a framework that emphasizes how the system produces variability, self-regulation, change, and inheritance will give rise to similar progress.

Analysis of some of the best specified cognitive-developmental change mechanisms provides reason to believe that this hope will be fulfilled. The models involve widely varying content domains and levels of analysis: synaptogenesis, associative competition, encoding, and analogical reasoning. They also have been used to explain developments that occur at very different ages: infancy, the toddler period, and later childhood. Despite these differences, the well-worked-out models all include processes for generating variation, self-regulation, adaptive change, and inheritance. Thus, there is reason to hypothesize that one unity among otherwise diverse cognitive-developmental change mechanisms is that they all serve these evolutionary functions.

3

Cognitive Variability: The Ubiquity of Multiplicity

No one doubts that immense variability exists at the neural level. Even when the identical stimulus is presented repeatedly within a single experimental session, the response of an individual neuron varies from trial to trial. Similarly, with low-level cognitive processes such as association, there is no disagreement concerning the existence of competing units. Models of associative memory, both symbolic (e.g., Gillilund & Shiffrin, 1984) and subsymbolic (e.g., Seidenberg & McClelland, 1990), are predicated on the assumptions that stimuli have multiple associations and that these varying associations influence the way in which we remember.

Higher level cognition, however, has been treated differently. Many models are universalist: Everyone is depicted as proceeding in the same way when relevant stimuli are presented. Other models are comparative; they hypothesize different ways of thinking among groups defined on the basis of such characteristics as age, expertise, or aptitudes, but hypothesize a single consistent kind of reasoning within each group. Thus, 8-year-olds might be depicted as performing in one way and 5-year-olds in another, experts in one way and novices in another, people with high spatial ability in one way and those with low spatial ability in another, and so on. The finest differentiations that are typically made within these comparative approaches examine individual differences within people of a single age; for example, reflective 8-year-olds are described as taking a long time but answering accurately on the Matching Familar Figures Test, and impulsive 8-year-olds are described as answering more quickly but less accurately (Kogan, 1983).

The main purpose of this chapter is to summarize the rapidly growing body of research suggesting that variability is actually a pervasive reality in high-level, as well as low-level, cognition. To place this work in context, however, it seems useful first to briefly consider some prominent examples of universalist and comparative models of cognition and then to consider why they might be proposed and widely accepted even if thinking is far more variable than they depict it as being.

Alternative Approaches to Cognitive Variability

Universalist Approaches

A great deal of cognitive research has been devoted to identifying *the* processing approach that people use on a particular task. This universalist approach has led to many influential models and theories. These include models of language processing (Carpenter & Just, 1975; Glucksberg & Keysar, 1990), spatial processing (Cooper & Shepard, 1973), inductive and deductive reasoning (Anderson, 1983; Guyote & Sternberg, 1981; Holyoak, Koh, & Nisbett, 1989; Johnson-Laird, Byrne, & Tabossi, 1989), judgment and decision making (Kahneman & Tversky, 1982; Tversky, Sattath, & Slovic, 1988), scientific reasoning (McCloskey, 1983; Proffitt, Kaiser, & Whelan, 1990), and many other types of tasks.

Such models are parsimonious and represent a logical place to start in trying to characterize the cognitive processes through which people draw inferences, reach decisions, reason, and solve problems. However, because the models are based on data averaged over subjects, they run the risk of not accurately reflecting what any individual is doing. Sidman (1952) and Estes (1956) were among the earliest to warn of this danger. They noted that continuous group learning curves can mask discontinuous individual learning. Even if each individual in a sample progresses from 0% to 100% correct performance in a single trial, the percent correct for the group can still take on the S-shape often used to infer that learning over trials is continuous (Figure 3.1). On the basis of this and similar evidence, Newell (1973) formulated his "Second Injunction of Psychological Experimentation": *Never average over methods*. In particular, he advocated careful experimental analyses of each subject's performance so that "we can settle what method he did indeed use" (p. 296).

The point made by Sidman, Estes, and Newell has been quite widely accepted, at least theoretically. I suspect that few researchers believe that all individuals perform higher level cognitive tasks in the same way, even if their models depict them as doing so. One response to this conflict has been to study, and sometimes model, between-subject variability in thinking.

Comparative Approaches

A common approach to between-subject variability is to specify two or more processing approaches that could be used on a task and to hypothesize that one group of people uses one approach and a different group a different approach. This *comparative paradigm* is a dominant strategy in studies of cognitive development—younger children are often said to think in one way, older children in another. The basic research strategy has become quite common among investigators in other areas of psychology as well.

Students of individual differences have been in the forefront of those documenting the use of different approaches by different people. In many such studies, people who are high in some ability are viewed as using one strategy and people lower in that ability another. For example, in several of the above-cited

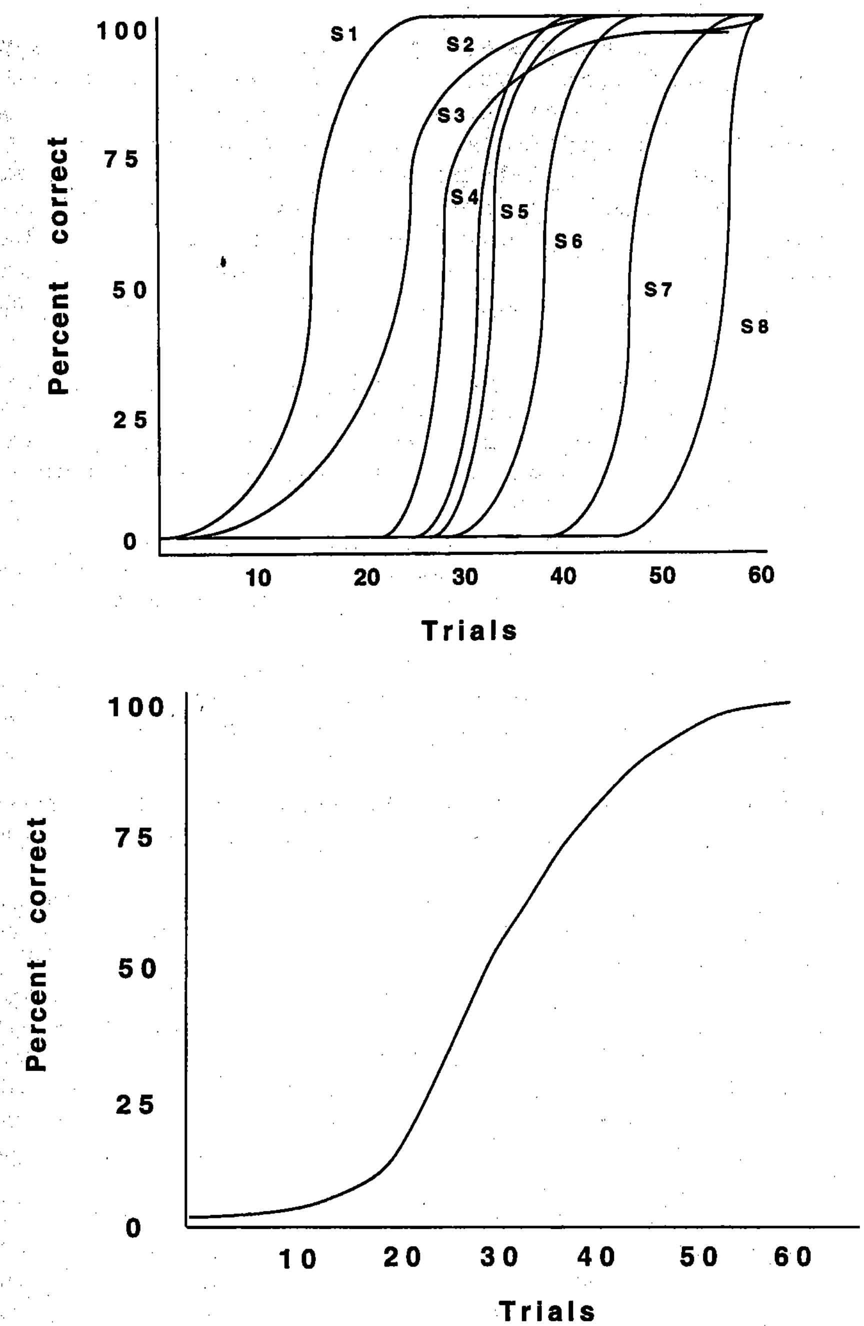

Figure 3.1. Continuous group learning curves can arise from averaging discontinuous individual learning patterns. Top graph shows eight hypothetical subjects, each of whom moved rapidly from inaccurate to accurate performance during a learning procedure. Bottom graph shows the continuous learning curve generated by averaging their performance.

domains in which initial models depicted everyone as using the same strategy, later models described people high in spatial ability as using spatially demanding strategies and people lower in spatial ability as using approaches with a greater emphasis on verbal processes. These include models of language comprehension (Mathews, Hunt, & McCleod, 1980), spatial problem solving (Cooper & Regan, 1982; Egan & Grimes-Farrow, 1982; Just & Carpenter, 1985; Kyllonen, Lohman, & Snow, 1984), and deductive reasoning (Sternberg & Weil, 1980). The findings and accompanying models from these studies raise fundamental doubts about the accuracy of the universalist one-size-fits-all models, as noted in the following comments:

> Discussion of the appropriate models for psycholinguistic tasks is usually couched in general terms (i.e., "What models apply to people?") Our results can be seen as a reminder that this approach is too simplistic. The same ostensibly linguistic task can be approached in radically different ways by different people. (MacLeod, Hunt, & Mathews, 1978, p. 506)
>
> The factor analysis methodology assumes that all subjects use the same general processes and structures on a test. This assumption is incorrect, and its violation may account for many of the confusions in the psychometric literature. (Just & Carpenter, 1985, p. 168)
>
> Too often psychologists set out to study *the* way that a task is performed, and miss one of the most interesting and general aspects of human cognitive performance: that there is more than one way to skin a cat. (Farah & Kosslyn, 1982, p. 164)

These empirical studies and theoretical arguments represent exactly the types of modification of traditional approaches envisioned in the critiques of Sidman, Estes, and Newell.

The same logic has been used in numerous other contexts. For example, a number of studies of expertise have contrasted the "expert strategy" with the "novice strategy." Mental calculation experts have been described as calculating from left to right and novices from right to left (Staszewski, 1988), expert physicists as representing problems in terms of qualitative characteristics and novices as representing them in terms of equations (Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980; Simon & Simon, 1978), expert headwaiters as formulating hierarchical representations with multiple levels for remembering restaurant orders and novices as representing the orders at a single level (Ericsson & Polson, 1988), and so on.

Many cross-cultural investigations also have followed this comparative paradigm. For example, schooled Moroccans have been said to rehearse on serial recall tasks but unschooled Moroccans not to rehearse (Wagner, 1978), Western children to ask questions to acquire information but West African children not to do so (Greenfield & Lave, 1982), and Australian aboriginal children to use spatial strategies but urban children verbal ones to recall the locations of objects (Kearins, 1981).

Similar logic is common within "noncognitive" areas, such as the study of personality. For example, mastery-oriented individuals are depicted as persisting in the face of adversity, whereas helpless individuals are said to give up (Dweck & Leggett, 1988), sex-typed individuals are depicted as organizing information in terms of gender whereas androgynous individuals are said to rely less on this categorization (Bem, 1981), and optimists are said to persist in attempts to reduce the discrepancies between their desired and actual self-images under conditions where pessimists would accept them (Carver & Scheier, 1992).

A variant of the comparative approach has been to distinguish between classes of situations rather than classes of people. For example, people have been characterized as using rational processes when not under time demands, but as taking shortcuts when under them (Payne, Bettman, & Johnson, 1993; Reder, 1987), as using spatial strategies when given instruction in visualization and algorithmic strategies when taught an algorithmic solution (Kyllonen, Lohman, & Snow, 1984; Sternberg & Ketron, 1982; Sternberg & Weil, 1980), and as employing systematic strategies under conditions of high involvement and heuristic ones under conditions of lower involvement (Chaiken, 1980).

Why Have Universalist and Comparative Approaches Been So Popular?

All of these models, both the universalist and the comparative ones, depict each individual as consistently behaving in a particular way within a particular situation. Their basic claim is: "This is what this class of people does when faced with this situation." One explanation for the pervasiveness of such depictions is that they reflect reality; perhaps individuals really do approach higher level cognitive tasks in very consistent ways. Another possibility, though, is that the models overstate the degree of consistency, and that there is considerable within-subject as well as between-subject variability, even under essentially identical conditions (i.e., same child, same problem, same task instructions, same experimental context, measures obtained close in time).

Why might psychological models depict people as thinking in very consistent ways if their thinking is actually quite variable? Three types of reasons seem likely: pragmatic, perceptual, and technological.

Pragmatic Reasons for Underestimating Cognitive Variability

Portraying cognition as simpler and less variable than it really is has a number of practical advantages. Models that say that a given individual will always proceed in a particular way are simpler and easier to test than models that recognize that people think in different ways on different occasions, even regarding identical problems. Another practical reason for understating cognitive variability is that widely used methods, such as chronometric analyses, tend to yield more orderly results when data are averaged over many trials. Recognizing that performance on different trials was generated via different processes raises complex data analysis issues. Related to these reasons, straightforward, easily tested models and use of standard data-analysis methods help in achieving personal and professional goals of investigators, such as publishing the research in prestigious journals, getting it described in textbooks, and enticing other researchers to extend the work.

Perceptual Reasons for Underestimating Cognitive Variability

Research on social perception suggests another reason why models, in particular comparative models, might depict members of an age, ability, personality, or cultural group as being more consistent than they really are: People are biased toward exaggerating between-group behavioral variability and minimizing within-group variability (Fiske & Taylor, 1991). Information that enhances between-group differences is especially likely to be encoded and remembered; information that indicates variability within individuals is especially likely to be forgotten or not learned in the first place (Krueger, Rothbart, & Siriam, 1989; Park & Hastie, 1987). Psychologists who attempt to describe groups defined by age, ability, personality, or culture are unlikely to be immune to these influences on how people perceive other people. Thus, selective perception and memory, as well as pragmatic factors related to personal advancement and to communication and testing of models, may predispose us to underestimate within-subject variability.

Technological Reasons for Underestimating Cognitive Variability

Research is limited by available technologies, as well as by the conceptual understanding of the researcher. Appreciation of the full degree of variability that is present in cognition depends on being able to accurately assess each subject's thinking on each trial. Until recently, this was almost impossible. Researchers needed either to settle for simple measures, such as percent correct or mean solution times, or for less formal, often unreliable, on-line observations of behavior.

In recent years, however, ability to obtain precise and reliable trial-by-trial assessments has been greatly enhanced by videocassette technology. This technology allows detailed observation of the audible and visible concomitants of higher level cognition. In so doing, it also allows more valid inferences about processes underlying the observable behavior than could be achieved without it. Thus, as often happens, technological innovations have led to scientific progress.

Prior to the advent of high quality, low cost videorecording systems, it was at best difficult to obtain reliable observations of the behavioral concomitants of higher level cognition. Often, it was impossible. This can be illustrated with regard to assessments of strategy use. Dedicated observers, notebooks in hand, could watch children and scribble impressions while trying not to miss anything. The limits of this method were obvious, though. In all but the simplest cases, part of the challenge is to classify the strategies. Ideally, strategy classification is an iterative process, in which hypothesized classification systems are revised as often as necessary to meet the twin goals of minimizing the number of categories and maximizing their fit to observed behavior. Without a permanent record of the behavior, however, such iterative revision is impossible; the experiment often will be over, and children's behavior gone, before a good classification system can be reached.

Videorecording provides a solution to this problem. We can try out alternative coding systems until we settle on one that seems best; then we can apply that system to all of the behavior that has been recorded. Videorecordings also allow

repeated scrutiny of behavior on particular trials, thus enhancing the accuracy of classification on each trial. Moreover, the technology can lead to breakthroughs in understanding, by allowing repeated analysis of especially revealing episodes.

Videorecording technology has proven especially useful in studying young children's thinking. Response methods that are widely used with adults, such as button boxes, keyboards, and voice-activated relays, cannot generally be used with infants, toddlers, or preschoolers. The children's lack of typing skills, tendency to forget which button goes with which response, and high frequency of unsolicited vocalizations generally rule out these methods. In contrast, by bringing together for each trial the overt behaviors that accompanied processing, the answer that was stated, the time it took to produce that answer, and the verbal description of strategy use, videorecording makes possible reliable and valid trial-by-trial strategy assessments for young children as well as older individuals.

To summarize, pragmatic, perceptual, and technological considerations have militated toward investigators portraying cognition as less variable than it really is. They have exercised this effect both directly, as described above, and indirectly, through their impact on the theories that became accepted as the major ones in the field and the tasks that became viewed as critical. Below I describe some of the experiences that have led to my own increasing conviction that cognitive variability is a pervasive and important phenomenon.

The Variability of Higher Level Cognition

I was not always so impressed with the variability of individual children's thinking. When first studying children's thinking, I thought that cognitive development was best described as a sequence of increasingly sophisticated rules for solving problems. I applied this approach to describing development of understanding of balance scales, projection of shadows, and fullness (Siegler, 1976; 1978; Siegler & Vago, 1978), conservation of liquid quantity, solid quantity, and number (Siegler, 1981), and time, speed, and distance (Siegler, 1983; Siegler & Richards, 1979).

These rule-based models painted a very orderly picture of development; in fact, they were as clear an example of the one-child:one-way-of-thinking approach as any that could be cited. The Figure 3.2 model of development of competence in solving balance scale problems illustrates this orderliness. In Rule I, children judge solely on the basis of the relative amounts of weight on the two sides of the fulcrum. In Rule II, they also consider relative distances from the fulcrum, but only when the weights on the two sides are equal. In Rule III, they generalize their consideration of distance to include all situations. Finally, in Rule IV, they proceed as in Rule III, unless one side has more weight and the other side more distance. In this case, they compute torques and choose the side with greater torque as the one that will go down. Thus, moving from Rule I to Rule II involves consideration of a new dimension, moving to Rule III involves generalizing consideration of this new dimension to all cases, and moving to Rule IV involves identifying a quantitative formula that solves all problems.

These rule-based descriptions fit the data well. The primary evidence came