1 *Prospects and limits of the empirical study of expertise: an introduction*

K. ANDERS ERICSSON AND JACQUI SMITH

Research on expertise may be one of the most rapidly expanding areas within cognitive psychology and cognitive science. Typically, when a topic becomes popular in psychology, the research approach and the methodology associated with it are also accepted, and the pressure to demonstrate the utility and feasibility of the approach diminishes. Efforts are directed instead toward the theoretical integration of research findings. Furthermore, popularity of a new approach nearly always means that many investigators will adopt it. An even larger number of investigators, however, will adopt only the terminology and will attempt to modify other research approaches to encompass the new concepts. That, in turn, leads to diffusion of the defining characteristics of the "new" approach, making straightforward attempts to integrate published research findings difficult. Because of this process of diffusion, often the new approach will no longer be readily distinguishable from previous alternative research approaches.

In this introductory chapter we attempt to provide a conceptual framework for distinguishing important characteristics of the *original expertise approach*. Our chapter consists of three sections. The first section attempts to characterize the study of expertise in the most general and domain-independent manner so that we can compare the expertise approach with a number of alternative approaches that had similar objectives. The focus of this section is on briefly reviewing some of the outcomes and failures of the earlier approaches. Our goal is to show that the expertise approach can account for these failures at the expense of greater empirical and theoretical complexity. In the second section we specify the nature of the original expertise approach and methodology. Here the pioneering work on chess expertise by de Groot (1978) and Chase and Simon (1973) is used to exemplify the sequence of research steps that characterized the original expertise approach. In the final section we elaborate criteria for these steps and use these criteria to discuss and review the prospects for, and limits of, more recent research on expertise.

DEFINITION OF OUTSTANDING PERFORMANCE AND EXPERTISE: A COMPARISON

On the most general level, the study of expertise seeks to understand and account for what distinguishes outstanding individuals in a domain from less outstanding individuals in that domain, as well as from people in general. We deliberately use the vague term "outstanding" because by not specifying more detailed criteria we are able to point to a number of distinctly different scientific approaches that have addressed the same problem.

In nearly all human endeavors there always appear to be some people who perform at a higher level than others, people who for some reason stand out from the majority. Depending on the historical period and the particular activity involved, such individuals have been labeled exceptional, superior, gifted, talented, specialist, expert, or even lucky. The label used to characterize them reflects an attribution of the major factor responsible for their outstanding behavior, whether it is intended to or not. Scientific efforts to understand the sources of such outstanding behavior have been guided by similar conceptions and attributions.

We limit our discussion to those cases in which the outstanding behavior can be attributed to relatively stable characteristics of the relevant individuals. We believe that stability of the individual characteristics is a necessary condition for any empirical approach seeking to account for the behavior with reference to characteristics of the individual. This constraint does not distinguish whether the characteristics are inherited or acquired. It does, however, eliminate a large number of achievements due to unique immediate environmental circumstances.

The most obvious achievements to be excluded by the stability constraint are those that involve events of fair games of chance, such as winning a large amount of money in a single lottery. More interestingly, the same criterion rules out achievements that occur only once in a lifetime, such as a single scientific discovery, a major artistic creation, a historically significant decision or prediction, or a single victory in a sport. This, of course, does not mean that we reject the possibility of defining criteria for outstanding performances in the arts, sciences, and sports arenas. It does mean, however, that a single achievement in a unique situation does not allow us to infer that the achievement was solely due to the particular individual's characteristics.

In order to support an attribution to the stable characteristics of a person, ideally one would require a series of outstanding achievements under different circumstances. Furthermore, one would like to have a larger group of other individuals (a "control" group of sorts) who have experienced similar opportunities to make contributions or to achieve. In the case in which many other individuals would be equally likely to achieve in similar situations, there is no need to attribute the achievement to special personal characteristics. Almost by definition the numbers of individuals given opportunities in some life realms to achieve and to stand out from the majority are small (e.g., heads

of state, army generals, people with vast economic resources). In such cases, even a stable series of achievements cannot unambiguously be linked to stable personal characteristics, because of the confounding influence of a unique stable situation.

Examination of our simple stable-characteristic constraint indicates that many achievements popularly acknowledged as evidence for expertise must be questioned and carefully scrutinized. Another important consequence of this constraint is more indirect and concerns the validity of social evaluation and perception of outstanding performance or ability. One would expect social evaluation to be greatly influenced by observations of previous performances (not all by the same individual) occurring under unique circumstances. A social judgment, then, might not be the most precise evaluation of an individual's current ability to perform. Ideally, one needs to determine the unique situation of the individual and to observe performances in standardized situations that allow interindividual comparisons (e.g., laboratory tasks or tests). Once it is possible to measure superior performance under standardized conditions, there is no need to rely on social indicators. Attuned to some of the difficulties of definition and assessment, let us now proceed to discuss some scientific approaches that have been directed toward accounting for outstanding or superior performance.

SCIENTIFIC APPROACHES TO ACCOUNTING FOR OUTSTANDING PERFORMANCE

Several different scientific approaches have been used to investigate outstanding performance. The constructs that have been investigated have primarily reflected popular attributions regarding the source of the outstanding behavior. These conceptualizations, in turn, have directly influenced what empirical evidence has been considered and collected. Table 1.1 summarizes the different types of stable personal characteristics that have been hypothesized to underlie outstanding performance and links those attributions to associated theoretical constructs and research methods. The attributed personal characteristics noted in Table 1.1 reflect a basic belief that behavior either is predominantly influenced by inherited qualities or is a function of learning and acquisition. Further, outstanding performance is attributed either to some general characteristic of the individual or to a specific aspect. The associated theoretical constructs and methodologies reflect these dimensions: *inherited* versus *acquired, general* versus *specific*. So, for example, the researcher will focus either on the effects of general traits (e.g., intelligence, personality), specific abilities (e.g., musical ability, spatial ability), and general life and educational experience (e.g., language, study strategies) or on domain-specific training and practice.

One's conception of the likely origins of outstanding performances will greatly influence the group of people selected for study, as well as the type of information sought concerning these individuals. For example, investigators

Table 1.1. *Different approaches to accounting for outstanding performance*

Attribution	Construct	Research approach
Primarily inherited		
General abilities	Intelligence, personality	Correlation with personality profile, general intelligence
Specific abilities	E.g., music ability, artistic ability, body build	Correlation with measures of specific ability
Primarily acquired		
General learning and experience	General knowledge and cognitive strategies	Investigation of common processing strategies
Domain-specific training and practice	Domain- or task-specific knowledge	Analysis of task performance, i.e., the expertise approach

pursuing an account in terms of general inherited capacities would be likely to consider individuals regardless of their domains and would be particularly interested in information allowing assessment of the genetic contribution. A longitudinal study of individuals identified as having exceptionally high intelligence, by Terman and his associates (Oden, 1968; Stanley, George, & Solano, 1977; Terman & Oden, 1947), illustrates this approach. A focus on domain-specific acquired characteristics would lead investigators to constrain themselves to one domain or task and to try to assess what was acquired (e.g., specific memory strategies), as well as the process of acquisition.

On a priori grounds one can argue that the most parsimonious theoretical account of outstanding performance is in terms of general, predominantly inherited characteristics. Indeed, in the history of scientific research on superior performance, that approach was initially preferred. It was primarily because of inability to explain certain empirical observations that accounts based on more specific abilities and acquired characteristics came to be seriously considered. We shall briefly consider some of those failures before turning to a consideration of the expertise approach that exemplifies the belief that specific acquired characteristics underlie outstanding performance.

Accounts in terms of general and specific inherited characteristics

If one wants to attribute outstanding performance to general inherited characteristics, it is reasonable to rely on readily available criteria to identify instances of outstanding behavior and of individuals who exhibit that behavior, criteria such as social evaluation and recognition by one's peers. In the first major study in that area, Galton (1869) used social recognition to identify eminent individuals in a wide range of fields and then studied their familial and

genetic origins. Galton argued that individuals gained eminence in the eyes of others because of a long-term history of achievement. Such achievement, he suggested, was the product of a blend of intellectual (natural) ability and personal motivation. He reported strong evidence for eminence's being limited to a relatively small number of families stemming from common ancestors, and he inferred that eminence was genetically determined.

Contemporary work in Galton's time and subsequent studies were directed at uncovering the loci of individual differences in general ability. The genetic nature of those general capacities led investigators to search for differences in basic characteristics of processes, such as the speed of mental processes as reflected by reaction time. In subsequent studies, however, individual differences in performance of simple tasks showed disappointingly low correlations, both among tasks and between performance and indices of ability, such as grade in school (Guilford, 1967).

More recent effort to uncover general basic cognitive processes that could account for individual differences have been inconclusive (Baron, 1978; Carroll, 1978; Cooper & Regan, 1982; Hunt, 1980). For example, research on individual differences in general memory ability has found low correlations of memory performance across different types of material and methods of testing, leading investigators to reject the idea of a general memory ability (Kelley, 1964). More direct evidence against stable basic memory processes comes from repeated demonstrations that memory performance for specific types of material can be drastically improved even after short periods of practice (Ericsson, 1985; Kliegl, Smith, & Baltes, 1989). Moreover, as Cooper and Regan noted (1982, p. 163), inadequacies in the definition and design of both cognitive tasks and intelligence measures create serious problems for interpreting correlations between measures of basic cognitive processes and ability.

Tests measuring general intelligence have been extremely useful for prediction and diagnosis in a wide range of situations, although there is considerable controversy about what they actually measure (Resnick, 1976; Sternberg, 1982). IQ tests, however, have been remarkably unsuccessful in accounting for individual differences in levels of performance in the arts and sciences and advanced professions, as measured by social indicators (e.g., money earned, status) and judgments (e.g., prizes, awards) (Tyler, 1965).

There were other lines of research that examined subjects with reliably superior performances and compared them with control groups. Much of that research was similarly motivated by the belief that exceptionally high levels of performance would reflect some basic exceptional ability involving attention (power or concentration), memory, general speed of reaction, or command of logic. Some investigators, however, focused on other stable individual characteristics, such as features of personality, motivation, and perceptual style (e.g., Cattell, 1963; Roe, 1953).

In the 1920s, three Russian professors examined the performance of eight grand masters (world-class chess players) on a wide range of laboratory tests for

basic cognitive and perceptual abilities (de Groot, 1946/1978). Surprisingly, the grand masters did not differ from control subjects in those basic abilities, but they were clearly superior in memory tests involving chess positions.

In the case of exceptional chess performance, superior spatial ability often is assumed to be essential (Chase & Simon, 1973; Holding, 1985). Doll and Mayr (1987) compared the performances of about thirty of the best chess players in what was then West Germany with those of almost ninety normal subjects of similar ages, using an IQ test with seven subscales. Only three of the subscales showed reliable differences, and somewhat surprisingly the largest difference between the two groups concerned higher scores for numeric calculation for the chess masters. Doll and Mayr (1987) found no evidence that chess players were selectively better on spatial tasks. In accounting for the unexpected superiority of the chess players on two of the subscales, Doll and Mayr (1987) argued that one reason could be that elite chess players had prior experience in coping with time pressure because of their past chess competitions. When the analysis was restricted to the group of elite chess players, none of the subscales of the IQ test was found to have a reliable correlation with chess-playing performance.

Of the research that has focused not on intelligence but on other relatively stable characteristics of individuals, that by Cattell (1963; Cattell & Drevdahl, 1955) is probably the best example. Cattell sought to determine whether the personality profiles for eminent researchers in physics, biology, and psychology could be distinguished from those of teachers and administrators in the same fields and from those of the general population. Compared with all other groups, top researchers were found to exhibit a consistent profile, being more self-sufficient, dominant, emotionally unstable, introverted, and reflective. Such a profile supports Galton's earlier opinion that eminence and outstanding achievement in a field are products not only of ability but also of aspects of personal motivation. Motivation and striving for excellence often are focused on a small number of domains or even a single domain, suggesting that aspects of motivation may well be acquired.

Despite these hints at possible personality patterns, the research approach of accounting for outstanding and superior performance in terms of general inherited characteristics has been largely unsuccessful in identifying strong and replicable relations. The search for links to specific inherited abilities has been similarly inconclusive. Indeed, as the specific characteristics proposed to account for the superior performance become integral to that performance, it becomes difficult to rule out the possibility that such characteristics have not been acquired as a result of many years of extensive training and practice. Investigators have therefore focused their attention on characteristics that appear in children and that reflect basic capacities for which a genetic origin is plausible. We shall briefly consider two examples of such basic capabilities, namely, absolute pitch among musicians and physiological differences among elite athletes.

A recent review of the research on absolute pitch shows that most of the

empirical evidence favors an account in terms of acquired skill (Ericsson & Faivre, 1988). The ability to recognize musical pitch is not an all-or-none skill, and many musicians have it to various degrees. They display the best performance on their own instruments, and their performance decreases as artificial tones from a tone generator are presented (Bachem, 1937). The ability to name pitches correctly is closely related to the amount of one's formal musical training (Oakes, 1955). Furthermore, pitch recognition can be dramatically improved with training, and one musician has documented how he acquired absolute pitch through long-term training (Brady, 1970).

Similarly, a recent review shows that many anatomical characteristics of elite athletes, such as larger hearts, more capillaries for muscles, and the proportions of different types of muscle fibers, are acquired during years of practice (Ericsson, 1990). Such findings showing the far-reaching effects of training do not, however, rule out possible genetic constraints. An individual's height and overall physique are determined by genetic factors (Wilson, 1986). Height and physique, for example, impose important constraints in many physical and sports domains, such as basketball, high jumping, gymnastics, ballet, and professional riding. It is also conceivable that genetic factors might influence the rate of improvement due to training. Nevertheless, training and preparation appear to be necessary prerequisites and important determinants of outstanding performance. We turn to a brief discussion of accounts of outstanding and superior performance based on acquired characteristics.

Accounts in terms of specific acquired characteristics: the expertise approach

In this brief review we have seen that the more parsimonious theoretical approaches relying on stable inherited characteristics seem inadequate to account for outstanding and superior performance. It is therefore necessary to consider accounts based on acquired characteristics. Here we need to identify not only what the acquired characteristics are but also the process by which they are acquired.

How long is the acquisition period, and over what time frame do we need to observe and monitor changes in performance? Simon and Chase (1973) were the first to observe that 10 years or more of full-time preparation are required to attain an international level of performance in chess. Studies by Hayes (1981) and Bloom (1985) revealed that a decade of intensive preparation is necessary to become an international performer in sports or in the arts or sciences. In a recent review, Ericsson and Crutcher (1990) found consistent support for the requirement of 10 years of intensive preparation in a wide range of studies of international levels of performance. Furthermore, Ericsson and Crutcher (1990) found for many domains that most international-level performers had been seriously involved in their domains before the age of 6 years. The period of preparation for superior performance appears to

cover a major proportion of these individuals' development during adolescence and early adulthood.

A detailed analysis of acquisition processes extending over decades under widely different environmental circumstances is extraordinarily difficult to conduct. Without a theoretical framework to outline the relevant aspects, the number of possible factors that could be critical to attain superior performance is vast. One can, of course, gain some idea of the range of factors by reading biographies and analyses of unusual events or circumstances in the lives of outstanding scientists and artists (Albert, 1983; McCurdy, 1983). It is unlikely, though, that descriptive studies seeking correlations between ultimate performance of individuals and information about their developmental histories will ever be able to yield conclusive results. A much more promising approach is offered by a careful analysis of the attained performance. This is the crux of the expertise approach.

The expertise approach differs from the approaches discussed earlier in some important respects. The other approaches were attempts to measure independently the constructs hypothesized to be the sources and bases of outstanding performance. In contrast, the expertise approach is an attempt to describe the critical performance under standardized conditions, to analyze it, and to identify the components of the performance that make it superior.

Two features distinguish the expertise approach from other approaches: first, the insistence that it is necessary to identify or design a collection of representative tasks to capture the relevant aspects of superior performance in a domain and to elicit superior performance under laboratory conditions; second, the proposal that systematic empirical analysis of the processes leading to the superior performance will allow assessment of critical mediating mechanisms. Moreover, it is possible to analyze the types of learning or adaptation processes by which these mechanisms can be acquired and to study their acquisition in real life or under laboratory conditions.

The expertise approach is more limited in its application than the other approaches reviewed earlier. Whereas the other approaches can use social indicators as criterion variables of outstanding performance, the expertise approach requires the design of a set of standardized tasks wherein the superior performance can be demonstrated and reliability reproduced. With this important limitation in mind, we now turn to a closer examination of the original expertise approach.

The original expertise approach: the pioneering work on chess

There is no consensus on how the expertise approach should be characterized. If one takes the original work on chess expertise by de Groot (1978) and Chase and Simon (1973), however, it is possible to extract three general characteristics. First, the focus is on producing and observing outstanding performance in the laboratory under relatively standardized conditions. Second, there is a theoretical concern to analyze and describe the cognitive

processes critical to the production of an outstanding performance on such tasks. Finally, the critical cognitive processes are examined, and explicit learning mechanisms are proposed to account for their acquisition.

If one is interested in reproducing superior performance under standardized conditions, one should give preference to domains in which there are accepted measures of performance. Chess provides such a domain. It is possible to measure an individual's chess-playing ability from the results of matches against different opponents in different tournaments (Elo, 1978). It is easy to select groups of chess players who differ sufficiently in chess ability that the probability of one of the weaker players beating one of the stronger players in a particular game is remote.

A critical issue in the expertise approach is *how to identify standardized tasks* that will allow the real-life outstanding performance to be reproduced in the laboratory. Because of the interactive nature of chess games and the vast number of possible sequences of moves, the same sequences of chess moves are hardly ever observed in two different chess games. Better chess players will consistently win over weaker chess players employing a wide variety of chess-playing styles. One could therefore argue that the better chess players consistently select moves as good as, or better than, the moves selected by weaker players. De Groot (1978) argued that it is possible to develop a collection of well-defined tasks capturing chess expertise by having chess players select the "best next move" for a number of different chess positions. Measurement of performance in this task requires that it be possible to evaluate qualitatively, on a priori grounds, the dependent variable, that is, the next chess move selected for a given chess position. It is not currently possible to evaluate the quality of chess moves for an arbitrary chess position. In fact, one international chess master claims to have spent a great part of his life unsuccessfully seeking to determine the best move for one particular chess position (Saariluoma, 1984).

De Groot (1978) collected think-aloud protocols from chess players of widely differing levels of expertise while they selected their best next moves for several chess positions. After extended analysis of these classic positions, however, he found that only *one* of them differentiated between grand masters and other chess experts who differed greatly in chess ability: All of the very best chess players selected better moves than did any of the comparatively weak players (nonoverlapping). Hence, he inferred that the task of selecting moves for that chess position must elicit cognitive processes that differentiate chess players at different levels of expertise.

Another pioneering aspect of de Groot's study was his use of verbal protocols. He was able to localize differences in cognitive processes between the grand masters and the other class experts by analyzing think-aloud protocols from his best-next-move task. He found that both masters and experts spent about 10 minutes before deciding on a move. In the beginning, the players familiarized themselves with the chess position, evaluated the position for strengths and weaknesses, and identified a range of promising moves. Later

they explored in greater depth the consequences of a few of those moves. On average, both masters and experts considered more than thirty move possibilities involving both Black and White and considered three or four distinctly different first moves.

De Groot (1978) first examined the possibility that, compared with chess experts, the grand masters were able to explore longer move combinations and thereby uncover the best move. He found, however, that the maximum depth of the search (i.e., the length of move combinations) was virtually the same for the two groups. When de Groot then focused his analysis on how the players came to consider different moves for the position, he did find differences. Few of the chess experts initially mentioned the best move, whereas most of the grand masters had noticed the best move during the familiarization with the position. More generally, de Groot argued, on the basis of his analysis of the protocols, that the grand masters perceived and recognized the characteristics of a chess position and evaluated possible moves by relying on their extensive experience rather than by uncovering those characteristics by calculation and evaluation of move possibilities. In some cases the discovery of promising chess moves was linked to the verbal report of a localized weakness in the opponent's chess position. Other grand masters discovered the same move without any verbal report of a mediating step (de Groot, 1978, p. 298). The superior chess-playing ability of more experienced chess players, according to de Groot, is attributable to their extensive experience, allowing retrieval of direct associations in memory between characteristics of chess positions and appropriate methods and moves. De Groot (1978, p. 316) argued that mastery in "the field of shoemaking, painting, building, [or] confectionary" is due to a similar accumulation of experiential linkings.

To examine the critical perceptual processing occurring at the initial presentation of a chess position, de Groot (1978) briefly showed subjects a middle-game chess position (2–10 seconds). Shortly after the end of the presentation the chess players gave retrospective reports on their thoughts and perceptions during the brief presentation and also recalled the presented chess position as best they could. From the verbal reports, de Groot found that the position was perceived in large complexes (e.g., a pawn structure, a castled position) and that unusual characteristics of the position (such as an exposed piece or a far-advanced pawn) were noticed. Within this brief time, the chess masters were found to integrate all the characteristics of the position into a single whole, whereas the less experienced players were not able to do so. The chess masters also often perceived the best move within that short exposure time. The analysis of the amount recalled from the various chess positions was consistent with the evidence derived from the verbal reports. Chess masters were able to recall the positions of all the 20–30 chess pieces virtually perfectly, whereas the positions recalled by the less experienced chess experts ranged from 50 to 70 percent.

The classic study of Chase and Simon (1973) followed up on this superior memory performance by chess masters for briefly presented chess positions.