CHAPTER 13

Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts' Performance on Representative Tasks

K. Anders Ericsson

The superior skills of experts, such as accomplished musicians and chess masters, can be amazing to most spectators. For example, club-level chess players are often puzzled by the chess moves of grandmasters and world champions. Similarly, many recreational athletes find it inconceivable that most other adults – regardless of the amount or type of training – have the potential ever to reach the performance levels of international competitors. Especially puzzling to philosophers and scientists has been the question of the extent to which expertise requires innate gifts versus specialized acquired skills and abilities.

One of the most widely used and simplest methods of gathering data on exceptional performance is to interview the experts themselves. But are experts always capable of describing their thoughts, their behaviors, and their strategies in a manner that would allow less-skilled individuals to understand how the experts do what they do, and perhaps also understand how they might reach expert level through appropriate training? To date, there has been considerable controversy over the extent to which experts are capable of explaining the nature and structure of their exceptional performance. Some pioneering scientists, such as Binet (1893/1966), questioned the validity of the experts' descriptions when they found that some experts gave reports inconsistent with those of other experts. To make matters worse, in those rare cases that allowed verification of the strategy by observing the performance, discrepancies were found between the reported strategies and the observations (Watson, 1913). Some of these discrepancies were explained, in part, by the hypothesis that some processes were not normally mediated by awareness/attention and that the mere act of engaging in self-observation (introspection) during performance changed the content of ongoing thought processes. These problems led most psychologists in first half of the 20th century to reject all types of introspective verbal reports as valid scientific evidence, and they focused almost exclusively on observable behavior (Boring, 1950).

In response to the problems with the careful introspective analysis of images and perceptions, investigators such as John B.

Watson (1920) and Karl Duncker (1945) introduced a new type of method to elicit verbal reports. The subjects were asked to "think aloud" and give immediate verbal expression to their thoughts while they were engaged in problem solving. In the main body of this chapter I will review evidence that this type of verbal expression of thoughts has not been shown to change the underlying structure of the thought processes and thus avoids the problem of reactivity, namely, where the act of generating the reports may change the cognitive processes that mediate the observed performance. In particular, I will describe the methods of protocol analysis where verbal reports are elicited, recorded, and encoded to yield valid data on the underlying thought processes (Ericsson & Simon 1980, 1984, 1993).

Although protocol analysis is generally accepted as providing valid verbalizations of thought processes (Simon & Kaplan, 1989), these verbal descriptions of thought sequences frequently do not contain sufficient detail about the mediating cognitive processes and the associated knowledge to satisfy many scientists. For example, these reports may not contain the detailed procedures that would allow cognitive scientists to build complete computer models that are capable of regenerating the observed performance on the studied tasks. Hence, investigators have continued to search for alternative types of verbal reports that generate more detailed descriptions. Frequently scientists require participants to explain their methods for solving tasks and to give detailed descriptions of various aspects. These alternative reporting methods elicit additional and more detailed information than is spontaneously verbalized during "think aloud." The desire for increased amounts of reported information is central to the study of expertise, so I will briefly discuss whether it is possible to increase the amount reported without inducing reactivity and change of performance. The main sections of this chapter describe the methods for eliciting and analyzing concurrent and retrospective verbal reports and how these methods have been applied to a number of domains of expertise, such as memory experts, chess masters, and medical specialists. The chapter is concluded with a broad overview of the issues of applying protocol analysis to the study of expert performance.

Historical Development of Verbal Reports on Thought Processes

Introspection or "looking inside" to uncover the structure of thinking and its mental images has a very long history in philosophy. Drawing on the review by Ericsson and Crutcher (1991), we see that Aristotle is generally given credit for the first systematic attempt to record and analyze the structure of sequences of thoughts. He recounted an example of series of thoughts mediating the recall of a specific piece of information from memory. Aristotle argued that thinking can be described as a sequence of thoughts, where the brief transition periods between consecutive thoughts do not contain any reportable information, and this has never been seriously challenged. However, such a simple description of thinking was not sufficiently detailed to answer the questions about the nature of thought raised by philosophers in the 17th, 18th, and 19th centuries (Ericsson & Crutcher, 1991).

Most of the introspective analysis of philosophers had been based on self-analysis of the individual philosophers' own thought. In the 19th century Sir Francis Galton along with others introduced several important innovations that set the groundwork for empirical studies of thinking. For example, Galton (1879, see Crovitz, 1970) noticed repeatedly that when he took the same walk through a part of London and looked at a given building on his path, this event triggered frequently the same or similar thoughts in memory. Galton recreated this phenomenon by listing the names of the major buildings and sights from his walk on cards and then presented a card at a time to observe the thoughts that were triggered. From this self-experiment Galton argued that thoughts reoccur with considerable frequency when the same stimulus is encountered.

Galton (1883) is particularly famous for the innovation of interviewing many people by sending out a list of questions about mental imagery – said to be the first questionnaire. He had been intrigued by reports of photographic memory and asked questions of the acuity of specific memories, such as the clarity and brightness of their memory for specific things such as their breakfast table. He found striking individual differences in the clarity or vividness, but no clear superiority of the eminent scientists; for example, Darwin reported having weak visual images. Now a hundred years later it is still unclear what these large individual differences in reported vividness of memory images really reflect. They seem almost completely unrelated to the accuracy of memory images and there is no reproducible evidence for individuals with photographic or eidetic memory (McKelvie, 1995; Richardson, 1988).

In one of the first published studies on memory and expertise Binet (1893/1966) reported a pioneering interview of chess players and their ability to play "blindfolded" without seeing a chess board. Based on anecdotes and his interviews Binet concluded that the ability required to maintain chess position in memory during blindfold play did not appear to reflect a basic memory capacity to store complex visual images, but a deeper understanding of the structure of chess. More troubling, Binet found that the verbal descriptions on the visual images of the mental chess positions differed markedly among blindfold chess players. Some claimed to see the board as clearly as if it were shown perceptually with all the details and even shadows. Other chess players reported seeing no visual images during blindfold play and claimed to rely on abstract characteristics of the chess position. Unfortunately, there was no independent evidence to support or question the validity of these diverse introspective reports. Binet's (1893/1966) classic report is a pioneering analysis of blindfold chess players' opinions and self-observations and illustrates the problems and limits of introspection.

In a similar manner Bryan and Harter (1899) interviewed two students of telegraphy as they improved their skill and found evidence for an extended plateau for both as they reached a rate of around 12 words per minute. Both reported that this arrest in development was associated with attempts to move away from encoding the Morse code into words and to encode the code into phrases. Subsequent research (Keller, 1958) has found that this plateau is not a necessary step toward expert levels of performance and referred to it as the phantom plateau.

In parallel with the interviews and the informal collection of self-observations of expertise in everyday life, laboratory scientists attempted to refine introspective methods to examine the structure of thinking. In the beginning of the 20th century, psychologists at the University of Würzburg presented highly trained introspective observers, with standardized questions and asked them to respond as fast as possible. After reporting their answers, the observers recalled as much as possible about the thoughts that they had while answering the questions. They tried to identify the most basic elements of their thoughts and images to give as detailed reports as possible. Most reported thoughts consisted of visual and auditory images, but some participants claimed to have experienced thoughts without any corresponding imagery – *imageless thoughts*. The principle investigator, Karl Bühler (1907), argued that the existence of imageless thoughts had far-reaching theoretical implications and was inconsistent with the basic assumption of Wilhelm Wundt (1897) that all thoughts were associated with particular neural activity in some part of the brain. Bühler's (1907) paper led to a heated exchange between Bühler's introspective observers, who claimed to have observed them, and Wundt (1907), who argued that these reports were artifacts of inappropriate reporting methods and the theoretical bias of the observers. A devastating methodological conclusion arose from this controversy: the existence of imageless thoughts could not be resolved empirically by the introspective method. This finding raised fundamental doubts about analytic introspection as a scientific method.

The resulting reaction to the crisis was to avoid the problem of having to trust the participants' verbal reports about internal events. Instead of asking individuals to describe the structure of their thoughts, participants were given objective tests of their memory and other abilities. More generally, experimental psychologists developed standardized tests with stimuli and instructions where the same pattern of performance could be replicated under controlled conditions. Furthermore, the focus of research moved away from complex mental processes, such as experts' thinking, and toward processes that were assumed to be unaffected by prior experience and knowledge. For example, participants were given well-defined simple tasks, such as memorization of lists of nonsense syllables, e.g., XOK, ZUT, where it is easy to measure objective performance. In addition, experimenters assumed that nonsense syllables were committed to memory without any reportable mediating thoughts, and the interest in collecting verbal reports from participants virtually disappeared until the cognitive revolution in the late 1950s.

In one of the pioneering attempts to apply this approach to the study of expertise, Djakow, Petrowski, and Rudik (1927) tested the basic abilities of world-class chess players and compared their abilities to other adults. Contrary to the assumed importance of superior basic cognitive ability and memory, the international players were only superior on a single test – a test involving memory for stimuli from their own domain of expertise, namely, chess positions. A few decades later de Groot (1946/1978d) replicated chess players' superior memory for chess positions and found that correct recall was closely related to the level of chess skill of the player.

Many investigators, including the famous behaviorist and critic of analytic introspection, John Watson, are very critical of the accuracy of verbal descriptions of skilled activities, such as where one looks during a golf swing (Watson, 1913). He realized that many types of complex cognitive processes, such as problem solving, corresponded to ongoing processes that were inherently complex and were mediated by reportable thoughts. In fact, Watson (1920) was the first investigator to publish a study where a participant was asked to "think aloud" while solving a problem. According to Watson, thinking was accompanied by covert neural activity of the speech apparatus that is frequently referred to as "inner speech." Hence, thinking aloud did not require observations by any hypothetical introspective capacity, and thinking aloud merely gives overt expression to these subvocal verbalizations. Many other investigators proposed similar types of instructions to give concurrent verbal expression of one's thoughts (see Ericsson & Simon, 1993, for a more extended historical review).

The emergence of computers in the 1950s and 1960s and the design of computer programs that could perform challenging cognitive tasks brought renewed interest in human cognition and higher-level cognitive processes. Investigators started studying how people solve problems and make decisions and attempted to describe and infer the thought processes that mediate performance. They proposed cognitive theories where strategies, concepts, and rules were central to human learning and problem solving (Miller, Galanter, & Pribram, 1960). Information-processing theories (Newell & Simon, 1972) sought computational models that could regenerate human performance on well-defined tasks by the application of explicit procedures. Much of the evidence for these complex mechanisms was derived from the researchers' own self-observation, informal interviews, and systematic questioning of participants.

Some investigators raised concerns almost immediately about the validity of these data. For example, Robert Gagné and his colleagues (Gagné & Smith, 1962) demonstrated that requiring participants to verbalize reasons for each move in the Tower of Hanoi improved performance by reducing the number of moves in the solutions and improving transfer to more difficult problems as compared to a silent control

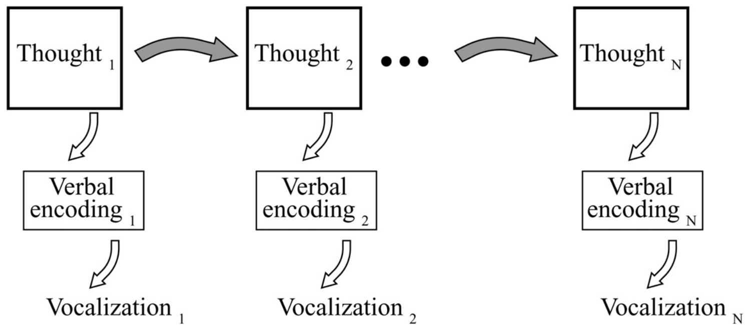

Figure 13.1. An illustration of the overt verbalizations of most thoughts passing through attention while a person thinks aloud during the performance of a task.

condition. Although improvements are welcome to educators, the requirement to explain must have changed the sequences of thoughts from those normally generated. Other investigators criticized the validity and accuracy of the retrospective verbal reports. For instance, Verplanck (1962) argued that participants reported that they relied on rules that were inconsistent with their observed selection behavior. Nisbett and Wilson (1977) reported several examples of experiments in social psychology, where participants gave explanations that were inconsistent with their observed behavior. These findings initially led many investigators to conclude that all types of verbal reports were tainted by similar methodological problems that had plagued introspection and led to its demise. Herb Simon and I showed in a review (Ericsson and Simon, 1980) that the methods and instructions used to elicit the verbal reports had a great influence on both the reactivity of the verbal reporting and on the accuracy of the reported information. We developed a parfticular type of methodology to instruct participants to elicit consistently valid nonreactive reports of their thoughts that I will describe in the next section.

Protocol Analysis: A Methodology for Eliciting Valid Data on Thinking

The central assumption of protocol analysis is that it is possible to instruct subjects to verbalize their thoughts in a manner that does not alter the sequence and content of thoughts mediating the completion of a task and therefore should reflect immediately available information during thinking.

Elicitation of Non-Reactive Verbal Reports of Thinking

Based on their theoretical analysis, Ericsson and Simon (1993) argued that the closest connection between actual thoughts and verbal reports is found when people verbalize thoughts that are spontaneously attended during task completion. In Figure 13.1 we illustrate how most thoughts are given a verbal expression.

When people are asked to think aloud (see Ericsson and Simon, 1993, for complete instructions), some of their verbalizations seem to correspond to merely vocalizing "inner speech," which would otherwise have remained inaudible. Nonverbal thoughts can also be often given verbal expression by brief labels and referents. Laboratory tasks studied by early cognitive scientists focused on how individuals applied knowledge and procedures to novel problems, such as mental multiplication. When, for example, one participant was asked to think aloud while mentally multiplying 36 by 24 on two test occasions one week apart, the following protocols were recorded:

> OK, 36 *times* 24, *um*, 4 *times* 6 *is* 24, 4, *carry the* 2, 4 *times* 3 *is* 12, 14, 144, 0, 2 *times* 6 *is* 12, 2, *carry the* 1, 2 *times* 3 *is* 6,

7, 720, 720, 144 *plus* 720, *so it would be* 4, 6, 864.

36 *times* 24, 4, *carry the – no wait,* 4, *carry the* 2, 14, 144, 0, 36 *times* 2 *is,* 12, 6, 72, 720 *plus* 144, 4, *uh, uh,* 6, 8, *uh,* 864.

In these two examples, the reported thoughts are not analyzed into their perceptual or imagery components as required by Bühler's (1907) rejected introspectionist procedures, but are merely vocalized inner speech and verbal expressions of intermediate steps, such as "carry the 1," "36," and "144 plus 720." Furthermore, participants were not asked to describe or explain how they solve these problems and do not generate such descriptions or explanations. Instead, they are asked to stay focused on generating a solution to the problem and thus only give verbal expression to those thoughts that spontaneously emerge in attention during the generation of the solution.

If the act of verbalizing participants' thought processes does not change the sequence of thoughts, then participants' task performance should not change as a result of thinking aloud. In a comprehensive review of dozens of studies, Ericsson and Simon (1993) found no evidence that the sequences of thoughts (accuracy of performance) changed when individuals thought aloud as they completed the tasks, compared to other individuals who completed the same tasks silently. However, some studies have shown that participants who think aloud take somewhat longer to complete the tasks – presumably due to the additional time required to produce the overt verbalization of the thoughts.

The same theoretical framework can also explain why other types of verbal-reporting procedures consistently change cognitive processes, like the findings of Gagné and Smith (1962). For example, when participants explain why they are selecting actions or carefully describe the structure and detailed content of their thoughts, they are not able to merely verbalize each thought as it emerges, they must engage in additional cognitive processes to generate the thoughts corresponding to the required explanations and descriptions. This additional cognitive activity required to generate the reports changes the sequence of generated thoughts (see Chi, Chapter 10, for a discussion of the differences between explanation and thinking aloud). Instructions to explain the reasons for one's problem solving and to describe the content of thought are reliably associated with changes in the accuracy of observed performance (Ericsson and Simon, 1993). Subsequent reviews have shown that the more recent work on effects of verbal overshadowing are consistent with reactive consequences of enforced generation of extensive verbal descriptions of brief experiences (Ericsson, 2002). Even instructions to generate self-explanations have been found to change (actually, improve) participants' comprehension, memory, and learning compared to merely thinking aloud during these activities (Ericsson, 1988a, 2003a; Neuman & Schwarz, 1998).

In summary, adults must already possess the necessary skills for verbalizing their thoughts concurrently, because they are able to think aloud without any systematic changes to their thought process after a brief instruction and familiarization in giving verbal reports (see Ericsson and Simon 1993, for detailed instructions and associated warm-up tasks recommended for laboratory research).

Validity of Verbalized Information while Thinking Aloud

The main purpose of instructing participants to give verbal reports on their thinking is to gain new information beyond what is available with more traditional measures of performance. If, on the other hand, verbal reports are the only source for some specific information about thinking, how can the accuracy of that information be validated? The standard approach for evaluating methodology is to apply the method in situations where other converging evidence is available and where the method's data can distinguish alternative models of task performance and disconfirm all but one reasonable alternatives.

Theories of human cognition (Anderson, 1983; Newell & Simon, 1972; Newell, 1990) proposed computational models that could reproduce the observable aspects of human performance on well-defined tasks through the application of explicit procedures. One of the principle methods applied by these scientists is an analysis of the cognitive task (see Chapter 11 by Schraagen for a discussion of the methods referred to as cognitive task analysis), and it serves a related purpose in the analysis of verbal protocols. Task analysis specifies the range of alternative procedures that people could reasonably use, in the light of their prior knowledge of facts and procedures, to generate correct answers to a task. Moreover, task analysis can be applied to the analysis of think-aloud protocols; for example, during a relatively skilled activity, namely, mental multiplication, most adults have only limited mathematical knowledge. They know the multiplication tables and only the standard "pencil and paper" procedure taught in school for solving multiplication problems. Accordingly, one can predict that they will solve a specific problem such as $36 \cdot 24$ by first calculating $4 \cdot 36 = 144$, then adding $20 \cdot 36 = 720$. More sophisticated adults may recognize that $24 \cdot 36$ can be transformed into $(30+6)(30-6)$ and that the formula $(a+b)(a-b) = a^2-b^2$ can be used to calculate $36 \cdot 24$ as $30^2-6^2 = 900-36 = 864$.

When adults perform tasks while thinking aloud the verbalized information must reflect information generated from the cognitive processes normally executed during the task. By analyzing this information, the verbalized sequences of thoughts can be compared to the sequence of intermediate results required to compute the answer by different strategies that are specified in a task analysis (Ericsson & Simon, 1993). The sequence of thoughts verbalized while multiplying $24 \cdot 36$ mentally (reproduced in the protocol examples above) agrees with the sequence of intermediate thoughts specified by one, and only one, of the possible strategies for calculating the answer.

However, the hypothesized sequence of intermediate products predicted from the task analysis may not perfectly correspond to the verbalizations. Inconsistencies may result from instances where, because of acquired skill, the original steps are either not generated or not attended as distinct steps. However, there is persuasive evidence for the validity of the thoughts that are verbalized, that is, that the verbalizations can reveal sequences of thoughts that match those specified by the task analysis (Ericsson & Simon, 1993). Even if a highly skilled participant's think-aloud report in the multiplication task only consisted of "144" and "720," the reported information would still be sufficient to reject many alternative strategies and skilled adaptations of them because these strategies do not involve the generation of both of the reported intermediate products. The most compelling evidence for the validity of the verbal reports comes from the use of task analysis to predict *a priori* a set of alternative sequences of concurrently verbalized thoughts that is associated with the generation of the correct answer to the presented problem.

Furthermore, verbal reports are only one indicator of the thought processes that occur during problem solving. Other indicators include reaction times (RTs), error rates, patterns of brain activation, and sequences of eye fixations. Given that each kind of empirical indicator can be separately recorded and analyzed, it is possible to examine the convergent validity established by independent analyses of different types of data. In their review, Ericsson and Simon (1993) found that longer RTs were associated with a longer sequence of intermediate reported thoughts. In addition, analyses show a close correspondence between participants' verbalized thoughts and the information that they looked at in their environment (see Ericsson & Simon, 1993, for a review).

Finally, the validity of verbally reported thought sequences depends on the time interval between the occurrence of a thought and its verbal report, where the highest validity is observed for concurrent, think-aloud verbalizations. For tasks with relatively short response latencies (less than 5 to 10 seconds), people are typically able to recall their sequences of thoughts accurately immediately after the completion of the task, and the validity of this type of retrospective reports remains very high. However, for cognitive processes of longer duration (longer than 10 to 30 seconds), recall of past specific thought sequences becomes more difficult, and people are increasingly tempted to infer what they must have thought, thus creating inferential biases in the reported information.

Other Types of Verbal Reports with Serious Validity Problems

Protocol analysis, as proposed by Ericsson and Simon (1980, 1984, 1993), specifies the constrained conditions necessary for valid, non-reactive verbalizations of thinking while performing a well-defined task. Many of the problems with verbally reported information obtained by other methods can be explained as violations of this recommended protocol-analysis methodology.

The first problem arises when the investigators ask participants to give more information beyond that which is contained in their recalled thought sequences. For example, some investigators ask participants *why* they responded in a certain manner. Participants may have deliberated on alternative methods; thus, their recalled thoughts during the solution will provide a sufficient answer, but typically the participants need to go beyond any retrievable memory of their processes to give an answer. Because participants can access only the end-products of their cognitive processes during perception and memory retrieval, and they cannot report why only one of several logically possible thoughts entered their attention, they must make inferences or confabulate answers to such questions.

In support of this type of confabulation, Nisbett and Wilson (1977) found that participants' responses to "why-questions" after responding in a task were in many circumstances as inaccurate as those given by other participants who merely observed these individuals' performance and tried to explain it without any memory or first-hand experience of the processes involved. More generally, Ericsson and Simon (1993) recommended that one should strive to understand these reactive, albeit typically beneficial, effects of instructing students to explain their performance. A detailed analysis of the different verbalizations elicited during "think-aloud" and "explain" instructions should allow investigators to identify those induced cognitive processes that are associated with changes (improvements) in their performance.

A very interesting development that capitalizes on the reactive effects of generating explanations involves instructing students to generate self-explanations while they read text or work on problems (Chi, de Leeuw, Chiu, & LaVancher, 1994; Renkl, 1997). Instructing participants to generate self-explanations has been shown to increase performance beyond that obtained with merely having them "think aloud," which did not differ from a control condition (Neuman, Leibowitz, & Schwarz, 2000). The systematic experimental comparison of instructions involving explanations or "thinking aloud" during problem solving has provided further insights into the differences between mechanisms underlying the generation of explanations that alter performance and those that merely give expression to thoughts while thinking aloud (Berardi-Coletta, Buyer, Dominowski, & Rellinger, 1995).

The second problem is that scientists are frequently primarily interested in the general strategies and methods participants use to solve a broad class of problems in a domain, such as mathematics or text comprehension. They often ask participants to describe their general methods after solving a long series of different tasks, which often leads to misleading summaries or after-the-fact reconstructions of what participants think they must have done. In the rare cases when participants have deliberately and consistently applied a single general strategy to solving the problems, they can answer such requests easily by recalling their thought sequence from any of the completed tasks. However, participants typically employ multiple strategies, and their strategy choices may change during the course of an experimental session. Under such circumstances participants would have great difficulty describing a *single* strategy that they used consistently throughout the experiment, thus their reports of such a strategy would be poorly related to their averaged performance. Hence, reviews of general strategy descriptions show that these reports are usually not valid, even when immediate retrospective verbal reports after the performance of each trial provide accounts of thought sequences that are consistent with other indicators of performance on the same trials (see Ericsson & Simon, 1993, for a review).

Similar problems have been encountered in interviews of experts (Hoffman, 1992). When experts are asked to describe their general methods in professional activities, they sometimes have difficulties, and there is frequently poor correspondence between the behavior of computer programs (expert systems) implementing their described methods and their observed detailed behavior when presented with the same tasks and specific situations. This finding has led many scientists studying expertise (Ericsson, 1996a; Ericsson & Lehmann, 1996; Ericsson & Smith, 1991; Starkes & Ericsson, 2003) to identify a collection of specific tasks that capture the essence of a given type of expertise. These tasks can then be presented under standardized conditions to experts and less-skilled individuals, while their think-aloud verbalizations and other process measures are recorded.

In sum, to obtain the most valid and complete trace of thought processes, scientists should strive to elicit laboratory conditions where participants perform tasks that are representative of the studied phenomenon and where verbalizations directly reflect the participants' spontaneous thoughts generated while completing the task. In the next section I will describe how protocol analysis has been applied to study experts' superior performance on tasks representative of their respective domain of expertise.

Protocol Analysis and the Expert-Performance Approach

The expert-performance approach to expertise (Ericsson, 1996a; Ericsson & Smith, 1991) examines the behavior of experts to identify situations with challenging task demands, where superior performance in these tasks captures the essence of expertise in the associated domain. These naturally emerging situations can be recreated as well-defined tasks calling for immediate action. The tasks associated with these situations can then be presented to individuals at all levels of skill, ranging from novice to international-level expert, under standardized conditions in which participants are instructed to give concurrent or retrospective reports.

In this section I will describe the expert-performance approach and illustrate its application of protocol analysis to study the structure of expert performance. First, de Groot's (1946/1978) pioneering work on the study of expert performance in chess will be described, followed by more recent extensions in the domain of chess as well as similar findings in other domains of expertise. Second, the issue of developing and validating theories of the mechanisms of individual experts will be addressed and several experimental analyses of expert performance will be described.

Capturing the Essence of Expertise and Analyzing Expert Performance

It is important to avoid the temptation to study differences in performance between experts and novices because there are readily available tasks to measure such differences. Researchers need to identify those naturally occurring activities that correspond to the essence of expertise in a domain (Ericsson, 2004, Chapter 38). For example, researchers need to study how chess players win tournament games rather than Just probing for superior knowledge of chess and test memory for chess games. Similarly, researchers need to study how doctors are able to treat patients with more successful outcomes rather than test their knowledge for medicine and memory of encountered patients. It is, however, difficult to compare different individuals' levels of naturally occurring performance in a domain because different individuals' tasks will differ in difficulty and many other aspects. For example, for medical doctors who primarily treat patients with severe and complex problems but with a relatively low frequency of full recovery, is their performance better than the performance of doctors who primarily treat patients with milder forms of the same disease with uniform recovery? Unless all doctors encounter patients with nearly identical conditions, it will be nearly impossible to compare the quality of their performance. The problem of comparing performers' performance for comparable tasks is a general challenge for measuring and capturing superior performance in most domains.

For example, chess players rarely, if ever, encounter the same chess positions during the middle part of chess games (Ericsson & Smith, 1991). Hence, there are no naturally occurring cases where many chess players select moves for the identical complex chess position such that the quality of their moves can be directly compared. In a path-braking research effort, de Groot (1946/1978) addressed this problem by identifying challenging situations (chess positions) in representative games that required immediate action, namely, the selection of the next move. De Groot then presented the same game situations to chess players of different skill levels and instructed them to think aloud while they selected the next chess move. Subsequent research has shown that this method of presenting representative situations and requiring generation of appropriate actions provides the best available measure of chess skill that predicts performance in chess tournaments (Ericsson, Patel, & Kintsch, 2000; van der Maas & Wagenmakers, 2005).

THE PIONEERING STUDIES OF CHESS EXPERTISE

In his pioneering research on chess expertise, de Groot (1946/1978) picked out chess positions that he had analyzed for a long time and established an informal task analysis. Based on this analysis he could evaluate the relative merits of different moves and encode the thoughts verbalized by chess players while they were selecting the best move for these positions.

The verbal protocols of both world-class and skilled club-level players showed that both types of players first familiarized themselves with the position and verbally reported salient and distinctive aspects of the position along with potential lines of attack or defense. The players then explored the consequences of longer move exchanges by planning alternatives and evaluating the resulting positions. During these searches the players would identify moves with the best prospects in order to select the single best move.

De Groot's (1946/1978) analysis of the protocols identified two important differences in cognitive processes that explained the ability of world-class players to select superior moves compared to club players. De Groot noticed that the less-skilled players didn't even verbally report thinking about the best move during their move selection, implying that they did not, in fact, think about it. Thus, their initial inferior representation of the position must not have revealed the value of lines of play starting with that move. In contrast, the world-class players reported many strong first moves even during their initial familiarization with the chess position. For example, they would notice weaknesses in the opponent's defense that suggested various lines of attack and then examine and systematically compare the consequences of various sequences of moves. During this second detailed phase of analysis, these world-class players would often discover new moves that were superior to all the previously generated ones.

MECHANISMS MEDIATING CHESS EXPERTISE

De Groot's analysis revealed two different mechanisms that mediate the world-class players' superiority in finding and selecting moves. The first difference concerns the best players' ability to rapidly perceive the relevant structure of the presented chess position, thus allowing them to identify weaknesses and associated lines of attack that the less-accomplished players never reported noticing in their verbal protocols. These processes involve rapid perception and encoding, and thus only the end products of these encoding processes are verbalized. There has been a great deal of research attempting to study the perceptual encoding processes by recording and analyzing eye fixations during brief exposures to reveal the cognitive processes mediating perception and memory of chess positions (see Gobet & Charness, Chapter 30). However, most of this research has not studied the task of selecting the best move but has used alternative task instructions, namely, to recall as many chess pieces as possible from briefly presented positions, or to find specific chess pieces in presented postions. These changes in the tasks appear to alter the mediating cognitive processes, and the results cannot therefore be directly integrated into accounts of the representative expert performance (Ericsson & Kintsch, 2000; Ericsson & Lehmann, 1996; Ericsson et al., 2000).

The second mechanism that underlies the superior performance of highly skilled players concerns a superior ability to generate potential moves by planning. De Groot's protocols showed that during this planning and evaluation process, the masters often discovered new moves that were better than those perceived initially during the familiarization phase. In a more recent study Charness (1981) collected think-aloud protocols on the planning process during the selection of a move for a chess position. Examples of an analysis of the protocols from a club-level and an expert-level chess player are given in Figure 13.2. Consistent with these examples, Charness (1981) found that the depth of planning increased with greater chess skill. In addition, there is evidence that an increase in the time available for planning increases the quality of the moves selected, where move selection during regular chess is superior to that of speed chess with its limited time for making the next move (Chabris & Hearst, 2003). Furthermore, highly skilled players have been shown to be superior in mentally planning out consequences of sequences of chess moves in experimental studies. In fact, chess masters, unlike less-skilled players, are able to play blindfold, without a visible board showing the current position, at a relatively high level (Chabris & Hearst, 2003; Karpov, 1995; Koltanowski, 1985). Experiments show that chess masters are able to mentally generate the chess positions associated with multiple chess games without any external memory support when the experimenter reads sequences of moves from multiple chess games (Saariluoma, 1991, 1995).

In sum, the analyses of the protocols along with experiments show that expert chess players' ability to generate better moves cannot be completely explained by their more extensive knowledge of chess patterns. Recognition of patterns and retrieval of appropriate moves that they have stored in memory during past experiences of chess playing is not sufficient to explain the observed reasoning abilities of highly skilled players. As their skill increases, they become increasingly able to encode and manipulate internal representations of chess positions to plan the consequences of chess moves, discover potential threats, and even develop new lines of attack (Ericsson & Kintsch, 1995; Saariluoma, 1992). (For a discussion of the relation between the superior memory for presented chess positions and the memory demands integral to selecting chess moves, see Ericsson et al., 2000, and Gobet & Charness, Chapter 30.)

MEDICINE AND OTHER DOMAINS

The expert-performance approach has been applied to a wide range of domains, where skilled and less-skilled performers solve representative problems while thinking aloud. When the review is restricted to studies in domains that show reproducibly superior performance of experts, the think-aloud protocols reveal patterns of reports that are consistent with those observed in chess. For example, when expert snooker players are instructed to make a shot for a

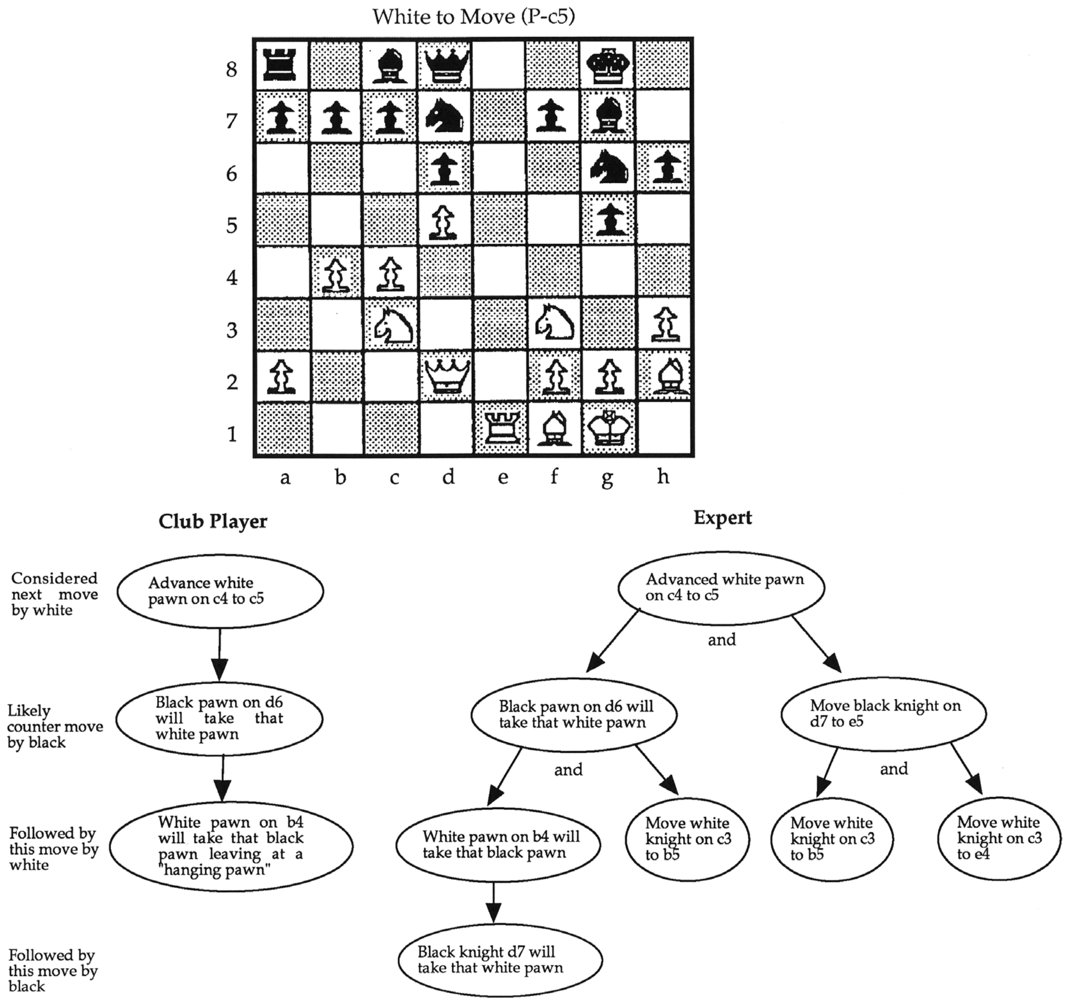

Figure 13.2. A chess position presented to chess players with the instruction to select the best next move by white (top panel). The think-aloud protocols of a good club player (chess rating = 1657) and a chess expert (chess rating = 2004) collected by Charness (1981) are shown in the bottom panel to illustrate differences in evaluation and planning for one specific move, P-c5 (white pawn is moved from c4 to c5), which is the best move for this position. Reported considerations for other potential moves have been omitted. The chess expert considers more alternative move sequences and some of them to a greater depth than the club player does. (From Ericsson, K. A., & Charness, N., 1994, Expert performance: Its structure and acquisition. *American Psychologist*, *49(8)*, 725–747, Figure 13.2 copyright American Psychological Association).

given configuration of pool balls, they verbalize deeper plans and more far-reaching exploration of consequences of their shots than less-skilled players (Abernethy, Neal, & Koning, 1994). Similarly, athletes at expert levels given protocols from dynamic situations in baseball (French, Nevett, Spurgeon, Graham, Rink, & McPherson, 1996) and soccer (Ward, Hodges, Williams, & Starkes, 2004) reveal a more complete and superior representation of the current game situation that allow them to prepare for future immediate actions better than less-skilled players in the same domains. In domains involving perceptual diagnosis, such as in the interpretation of Electrocardiograms (ECG) (Simpson & Gilhooly, 1997) and microscopic pathology (Crowley, Naus, Stewart,

& Friedman, 2003), verbal protocols reveal that the experts are able to encode essential information more accurately and are more able to integrate the information into an accurate diagnosis.

Most of the research on medical diagnosis has tried to minimize the influence of perceptual factors and has relied primarily on verbal descriptions of scenarios and patients. This research on medical expertise has shown that the process of generating a diagnosis becomes more efficient as medical students complete more of their medical training. The increase in efficiency is mediated by higher levels of representation that is acquired to support clinical reasoning (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1993). When studies present very challenging medical problems to specialists and medical students, the experts give more accurate diagnoses (Ericsson, 2004; Norman, Trott, Brooks, & Smith, 1994). The specialists are also more able to give complete and logically supported diagnoses (Patel & Groen, 1991) that appear to reflect higher-level representations that they have acquired to support reasoning about clinical alternative diagnoses (Ericsson & Kintsch, 1995; Ericsson et al., 2000; Patel, Arocha, & Kaufmann, 1994).

There are also studies showing differences in knowledge between experts and less-accomplished individuals that mediate successful task performance in experimental design of experiments in psychology (Schraagen, 1993) and detection of fraud in financial accounting (Johnson, Karim, & Berryman, 1991). The work on accounting fraud was later developed into a general theory of fraud detection (Johnson, Grazioli, Jamal, & Berryman, 2001). In this handbook there are discussions of the applications of verbal report methodology to study thinking in several different domains of expertise, such as medicine (Norman, Eva, Brooks, & Hamstra, Chapter 19), software design (Sonnentag, Niessen, & Volmer, Chapter 21), professional writing (Kellogg, Chapter 22), artistic performance (Noice & Noice, Cahpter 28), chess playing (Gobet & Charness, Chapter 30), exceptional memory (Wilding & Valentine, Chapter 31), mathematical expertise (Butterworth, Chapter 32), and historical expertise (Voss & Wiley, Chapter 33).

The evidence reviewed in this section has been based primarily on findings that are based on averages across groups of experts. In the next section we will search for evidence on the validity of reported thoughts of individual experts as well as individual differences between different experts.

Individual Differences and Validity of Verbal Reports from Expert Performance

It is well established that to be successful in competitions at the international level, experts need to have engaged in at least ten years of intensive training – a finding that applies even to the most "talented" individuals (Ericsson Krampe, & Tesch-Romer, 1993; Simon & Chase, 1973). Consequently, researchers have not been surprised that verbal reports of experts and, thus, the corresponding sequences of reported thoughts, differ between expert performers – at least at the level of detailed thoughts. In the previous section I showed how protocols uncover many higher-level characteristics of expert performers' mediating mechanisms, such as skills supporting the expanded working memory (Ericsson & Kintsch, 1995). In this section I will discuss attempts to experimentally validate the detailed structure of the reported cognitive processes of individual expert performers.

The complexity of the knowledge and acquired skills of expert performers in most domains, such as chess and medicine, makes it virtually impossible to describe the complete structure of the expertise of an individual expert. For example, Allen Newell (personal communication) described a project in which one of his graduate students in the 1970s tried to elicit all the relevant knowledge of a stamp collector. After some forty hours of interviews, Newell and his student gave up, as there was no sight of the end of the knowledge that the expert had acquired. As it may be difficult, perhaps impossible, to describe all the knowledge and skills of experts, scientists should follow the recommendations of the expert-performance approach. Namely, they should focus on the reproducible structure of the experts' mechanisms that mediate their superior performance on representative tasks (Ericsson, 1996b). Consequently, I will focus on selected domains of expertise in which regularities in the verbal reports of different trials with representative tasks have been analyzed.

In the early applications of protocol analysis there were several studies that collected protocols from experts solving representative problems while thinking aloud. For example, Clarkson and Metzler (1960) collected protocols from a professional investor constructing portfolios of investments. Similar detailed analyses of individual experts from different domains have been briefly described in Ericsson and Simon (1993) and Hoffman (1992). These analyses were not, however, formally evaluated, and the proposed mechanisms were not demonstrated to account for reproducibly superior performance on representative tasks.

The most extensive applications of the expert-performance approach using protocol analysis to study individual experts have examined people with exceptional memory (Ericsson & Lehmann, 1996). In the introduction of this chapter I mentioned Binet's (1894) pioneering work studying individuals with exceptional memory for numbers. Several subsequent studies interviewed people with exceptional memory, such as Luria's (1968) Subject S and Hunt and Love's (1972) VP (see Wilding and Valentine, 1997, for a review). However, the first study to trace the development of exceptional memory from average performance to the best memory performance in the world (in some memory tasks) was conducted in a training study by Chase and Ericsson (1981, 1982; Ericsson, Chase, & Faloon, 1980). We studied a college student (SF) whose initial immediate memory for rapidly presented digits was around 7, in correspondence with the typical average (Miller, 1956), but he eventually acquired exceptional performance for immediate memory and after 200 hours of practice was able to recall over 80 digits in the digit-span task. During this extended training period SF gave retrospective reports on his thought processes after most memory trials. As his memory performance started to increase he reported segmenting the presented lists into 3-digit groups and, whenever possible, encoding them as running times for various races because SF was an avid cross-country runner. For example, SF would encode 358 as a very fast mile time, 3 minutes and 58 seconds, just below the 4-minute mile. The central question concerning verbal reports is whether we can trust the validity of these reports and whether the ability to generate mnemonic running-time encodings influences memory.

To address that issue Bill Chase and I designed an experiment to test the effects of mnemonic encodings and presented SF with special types of lists of constrained digits. In addition to a list of random digits we presented other lists that were constructed to contain only 3-digits groups that could not be encoded as running times, such as 364 as three minutes and sixty four seconds, in a list (364 895 481 . . .). As predicted his performance decreased reliably. In another experiment we designed digit sequences where all 3-digit groups could be encoded as running times (412 637 524 . . .) with a reliable increase in his associated performance. In over a dozen specially designed experiments it was possible to validate numerous aspects of SF's acquired memory skill (Chase & Ericsson, 1981, 1982; Ericsson, 1988b). Other investigators, such as Wenger and Payne (1995), have also relied on protocol analysis and other process-tracing data to assess the mechanisms of individuals who increased their memory performance dramatically with practice on a list-learning task.

More generally, this method has been extended to any individual with exceptional memory performance. During the first step, the exceptional individuals are given memory tasks where they could exhibit their exceptional performance while giving concurrent and/or retrospective verbal reports. These reports are then analyzed to identify the mediating encoding and retrieval mechanisms of each exceptional individual. The validity of these accounts is then evaluated experimentally by presenting each individual with specially designed memory tasks that would predictably reduce that individuals' memory performance in a decisive manner (Ericsson, 1985, 1988b; Wilding & Valentine, 1998). With this methodology, verbal reported mechanisms of superior performance have been validated with designed experiments in a wide range of domains, such as a waiter with superior memory for dinner orders (Ericsson & Polson, 1988a, 1988b), mental calculators (Chase & Ericsson, 1982) and other individuals with exceptional memory performance (Ericsson, 2003b; Ericsson, Delaney, Weaver, & Mahadevan, 2004).

Exceptional memory performance for numbers and other types of "arbitrary" information appears to require that the expert performers sustain attention during the presentation (Ericsson, 2003b). The difficulty to automate memory skills for encoding new stimuli makes this type of performance particularly amenable to examination with protocol analysis. More generally, when individuals change and improve their performance they appear able to verbalize their thought processes during learning (Ericsson & Simon, 1993). This has been seen to extend to learning of experts and their ability to alter their performance through deliberate practice (Ericsson et al., 1993). There is now an emerging body of research that examines the microstructure of this type of training and how additional specific deliberate practice improves particular aspects of the target performance in music (Chaffin & Imreh, 1997; Nielsen, 1999) and in sports (Deakin & Cobley, 2003; Ericsson, 2003c; Ward et al., 2004) – for a more extended discussion see the chapter by Ericsson (Chapter 38) on deliberate practice.

Conclusion

Protocol analysis of thoughts verbalized during the experts' superior performance on representative tasks offers an alternative to the problematic methods of directed questioning and introspection. The think-aloud model of verbalization of thoughts has been accepted as a useful foundation for dealing with the problems of introspection (see the entry on "Psychology of Introspection" in the *Routledge Encyclopedia of Philosophy* by Von Eckardt, 1998, and entries on "Protocol Analysis" in the *Companion to Cognitive Science* [Ericsson, 1998] and the *International Encyclopedia of the Social and Behavioral Sciences* [Ericsson, 2001]. This same theoretical framework for collecting verbal reports has led to the accumulation of evidence that has led many behaviorists to accept data on cognitive constructs, such as memory and rules (Austin & Delaney, 1998). Consequently, the method of protocol analysis provides a tool that allows researchers to identify information that pass through expert performers' attention while they generate their behavior without the need to embrace any controversial theoretical assumptions. In support of this claim, protocol analysis has emerged as a practical tool to diagnose thinking outside of traditional cognitive psychology and cognitive science. For example, designers of surveys (Sudman, Bradburn, & Schwarz, 1996), researchers on second-language learning(Green, 1998) and text comprehension passages (Ericsson, 1988a; Pressley & Afflerbach, 1995), and computer software developers (Henderson, Smith, Podd, & Varela-Alvarez, 1995; Hughes & Parkes, 2003) regularly collect verbal reports and rely on protocol analysis.

The complexity and diversity of the mechanisms mediating skilled and expert performance is intimidating. To meet these challenges it is essential to develop methods to allow investigators to reproduce the experts' superior performance under controlled and experimental conditions on tasks that capture the essence of expertise in a given domain. Process tracing, in particular protocol analysis, will be required to uncover detailed information about most of the important mechanisms that are responsible for the superiority of the experts' achievement. Only then will it be possible to discover their structure and study their development and refinement with training and deliberate practice.

References

Abernethy, B., Neal, R. J., & Koning, P. (1994). Visual-perceptual and cognitive differences between expert, intermediate, and novice snooker players. *Applied Cognitive Psychology*, *18*, 185–211.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Austin, J., & Delaney, P. F. (1998). Protocol analysis as a tool for behavior analysis. *Analysis of Verbal Behavior*, *15*, 41–56.

Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L., & Rellinger, E. R. (1995). Metacognition and problem solving: A process-oriented approach. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 205–223.

Binet, A. (1893/1966). Mnemonic virtuosity: A study of chess players. (Original paper appeared in 1893 and was translated by M. L. Simmel & S. B. Barron). *Genetic Psychology Monographs*, *74*, 127–162.

Binet, A. (1894). *Psychologie des grands calculateurs et joueurs d'echecs* [The psychology of great calculators and chess players]. Paris: Libraire Hachette.

Boring, E. B. (1950). *A history of experimental psychology*. New York: Appleton-Century Crofts.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, *16*, 153–184.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, *6*, 345–375.

Bühler, K. (1907). Tatsachen und Probleme zu einer Psychologie der Denkvorgaenge: I. Ueber Gedanken [Facts and problems in a psychology of thinking: I. On thoughts]. *Archiv für die gesamte Psychologie*, *9*, 297–365.

Chabris, C. F., & Hearst, E. S. (2003). Visualization, pattern recognition, and forward search: Effects of playing speed and sight of the position on grandmaster chess errors. *Cognitive Science*, *27*. 637–648.

Chaffin, R., & Imreh, G. (1997). "Pulling teeth and torture": Musical memory and problem solving. *Thinking and Reasoning*, *3*, 315–336.

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 467–476.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 16 (pp. 1–58). New York: Academic Press.

Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477.

Clarkson, G. P., & Metzler, A. H. (1960). Portfolio selection: A heuristic approach. *Journal of Finance*, *15*, 465–480.

Crovitz, H. F. (1970). *Galton's walk: Methods for the analysis of thinking, intelligence, and creativity*. New York: Harper & Row Publishers.

Crowley, R. S., Naus, G. J., Stewart, J., & Friedman, C. P. (2003). Development of visual diagnostic expertise in pathology: An information-processing study. *Journal of the American Medical Informatics Association*, *10*, 39–51.

Crutcher, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science*, *5*, 241–244.

de Groot, A. (1978). *Thought and choice and chess*. The Hague: Mouton. (Original work published 1946).

Deakin, J. M., & Cobley, S. (2003). A search for deliberate practice: An examination of the practice environments in figure skating and volleyball. In J. Starkes & K. A. Ericsson (Eds.), *Expert performance in sport: Recent advances in research on sport expertise* (pp. 115–135). Champaign, IL: Human Kinetics.

Djakow, J. N., Petrowski, N. W., & Rudik, P. A. (1927). *Psychologie des Schachspiels* [The psychology of chess]. Berlin: Walter de Gruyter.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58(5, Whole No. 270).

Ericsson, K. A. (1985). Memory skill. *Canadian Journal of Psychology*, *39*, 188–231.

Ericsson, K. A. (1988a). Concurrent verbal reports on reading and text comprehension. *Text*, *8*, 295–325.

Ericsson, K. A. (1988b). Analysis of memory performance in terms of memory skill. In R. J. Sternberg (Ed.), *Advances in the psychology of* *human intelligence*, Vol. 4 (pp. 137–179). Hillsdale, NJ: Erlbaum.

Ericsson, K. A. (Ed.) (1996a). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*. Mahwah, NJ: Erlbaum.

Ericsson, K. A. (1996b). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 1–50). Mahwah, NJ: Erlbaum.

Ericsson, K. A. (1998) Protocol analysis. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 425–432). Cambridge, MA: Basil Blackwell.

Ericsson, K. A. (2001). Protocol analysis in psychology. In N. Smelser & P. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 12256–12262). Oxford, UK: Elsevier.

Ericsson, K. A. (2002). Toward a procedure for eliciting verbal expression of nonverbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. *Applied Cognitive Psychology*, *16*, 981–987.

Ericsson, K. A. (2003a). Valid and non-reactive verbalization of thoughts during performance of tasks: Toward a solution to the central problems of introspection as a source of scientific data. *Journal of Consciousness Studies*, *10*, 1–18.

Ericsson, K. A. (2003b). Exceptional memorizers: Made, not born. *Trends in Cognitive Sciences*, *7*, 233–235.

Ericsson, K. A. (2003c). The development of elite performance and deliberate practice: An update from the perspective of the expert-performance approach. In J. Starkes & K. A. Ericsson (Eds.), *Expert performance in sport: Recent advances in research on sport expertise* (pp. 49–81). Champaign, IL: Human Kinetics.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, *10*, S1–S12.

Ericsson, K. A., Chase, W., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182.

Ericsson, K. A., & Crutcher, R. J. (1991). Introspection and verbal reports on cognitive processes – two approaches to the study of thought processes: A response to Howe. *New Ideas in Psychology*, *9*, 57–71.

Ericsson, K. A., Delaney, P. F., Weaver, G., & Mahadevan, R. (2004). Uncovering the structure of a memorist's superior "basic" memory capacity. *Cognitive Psychology*, *49*, 191–237.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Ericsson, K. A., & Kintsch, W. (2000). Shortcomings of article retrieval structures with slots of the type that Gobet (1993) proposed and modeled. *British Journal of Psychology*, *91*, 571–588.

Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence on maximal adaptations on task constraints. *Annual Review of Psychology*, *47*, 273–305.

Ericsson, K. A., Patel, V. L., & Kintsch, W. (2000). How experts' adaptations to representative task demands account for the expertise effect in memory recall: Comment on Vicente and Wang (1998). *Psychological Review*, *107*, 578–592.

Ericsson, K. A., & Polson, P. G. (1988a). Memory for restaurant orders. In M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 23–70). Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Polson, P. G. (1988b). An experimental analysis of a memory skill for dinner orders. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 305–316.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Bradford books/MIT Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (revised edition)*. Cambridge, MA: Bradford books/MIT Press.

Ericsson, K. A., & Smith, J. (1991). Prospects and limits in the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 1–38). Cambridge: Cambridge University Press.

French, K. E., Nevett, M. E., Spurgeon, J. H., Graham, K. C., Rink, J. E., & McPherson, S. L. (1996). Knowledge representation and problem solution in expert and novice youth baseball players. *Research Quarterly for Exercise and Sport*, 67, 386–395.

Gagné, R. H., & Smith, E. C. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology*, 63, 12–18.

Galton, F. (1879). Psychometric experiments. *Brain*, 2, 148–162.

Galton, F. (1883). *Inquiries into human faculty and its development*. New York: Dutton.

Green, A. J. F. (1998). *Using verbal protocols in language testing research: A handbook*. Cambridge, UK: Cambridge University Press.

Henderson, R. D., Smith, M. C., Podd, J., & Varela-Alvarez, H. (1995). A comparison of the four prominent user-based methods for evaluating the usability of computer software. *Ergonomics*, 39, 2030–2044.

Hoffman, R. R. (Ed.) (1992). *The psychology of expertise: Cognitive research and empirical AI*. New York: Springer-Verlag.

Hughes, J., & Parkes, S. (2003). Trends in the use of verbal protocol analysis in software engineering research. *Behaviour & Information Technology*, 22, 127–140.

Hunt, E., & Love, T. (1972). How good can memory be? In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 237–260). New York: Holt.

Johnson, P. E., Karim, J., & Berryman, R. G. (1991). Effects of framing on auditor decisions. *Organizational Behavior & Human Decision Processes*, 50, 75–105

Johnson, P. E., Grazioli, S., Jamal, K., & Berryman, R. G. (2001). Detecting deception: Adversarial problem solving in a low base-rate world. *Cognitive Science*, 25, 355–392.

Karpov, A. (1995). Grandmaster musings. *Chess Life*, November, pp. 32–33.

Keller, F. S. (1958). The phantom plateau. *Journal of the Experimental Analysis of Behavior*, 1, 1–13.

Koltanowski, G. (1985). *In the dark*. Coraopolis, PA: Chess Enterprises.

Luria, A. R. (1968). *The mind of a mnemonist*. New York: Avon.

McKelvie, S. J. (1995). The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery*, 19, 197–252.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review*, 63, 81–97.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart, and Winston.

Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? The case of analogical problem-solving. *British Journal of Educational Psychology*, 68, 15–24.

Neuman, Y., Leibowitz, L., & Schwarz, B. (2000). Patterns of verbal mediation during problem solving: A sequential analysis of self-explanation. *Journal of Experimental Education*, 68, 197–213.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nielsen, S. (1999). Regulation of learning strategies during practice: A case study of a single church organ student preparing a particular work for a concert performance. *Psychology of Music*, 27, 218–229.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.

Norman, G. R., Trott, A. D., Brooks, L. R., & Smith, E. K. M. (1994). Cognitive differences in clinical reasoning related to postgraduate training. *Teaching and Learning in Medicine*, 6, 114–120.

Patel, V. L., Arocha, J. F., & Kaufmann, D. R. (1994). Diagnostic reasoning and medical expertise. In D. Medin (Ed.), *The psychology of learning and motivation*, Vol. 30 (pp. 187–251). New York: Academic Press.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 93–125). Cambridge, MA: Cambridge University Press.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, 21, 1–29.

Richardson, J. T. E. (1988). Vividness and unvividness: Reliability, consistency, and validity of subjective imagery ratings. *Journal of Mental Imagery*, 12, 115–122.

Saariluoma, P. (1991). Aspects of skilled imagery in blindfold chess. *Acta Psychologica*, *77*, 65–89.

Saariluoma, P. (1992). Error in chess: The apperception-restructuring view. *Psychological Research/Psychologische Forschung*, *54*, 17–26.

Saariluoma, P. (1995). *Chess players' thinking*. London: Routledge.

Schmidt, H. G., & Boshuizen, H. (1993). On acquiring expertise in medicine. *Educational Psychology Review*, *5*, 205–221.

Schraagen, J. M. (1993). How experts solve a novel problem in experimental design. *Cognitive Science*, *17*, 285–309.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, *61*, 394–403.

Simon, H. A., & Kaplan, C. A. (1989). Foundations of cognitive science. In M. J. Posner (Ed.), *Foundations of cognitive science* (pp. 1–47). Cambridge, MA: MIT Press.

Simpson, S. A., & Gilhooly, K. J. (1997). Diagnostic thinking processes: Evidence from a constructive interaction study of electrocardiogram (ECG) interpretation. *Applied Cognitive Psychology*, *11*, 543–554.

Starkes, J., & Ericsson, K. A. (Eds.) (2003). *Expert performance in sport: Recent advances in research on sport expertise*. Champaign, IL: Human Kinetics.

Sudman, S., Bradburn, N. M., & Schwarz, N. (Eds.) (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

van der Maas, H. L. J., & Wagenmakers, E. J. (2005). A psychometric analysis of chess expertise. *American Journal of Psychology*, *118*, 29–60.

Verplanck, W. S. (1962). Unaware of where's awareness: Some verbal operants-notates, moments and notants. In C. W. Eriksen (Ed.), *Behavior and awareness – a symposium of research and interpretations* (pp. 130–158). Durham, NC: Duke University Press.

Von Eckardt, B. (1998). Psychology of introspection. In E. Craig (Ed.), *Routledge encyclopedia of philosophy* (pp. 842–846). London: Routledge.

Ward. P., Hodges, N. J., Williams, A. M., & Starkes, J. L. (2004). Deliberate practice and expert performance: Defining the path to excellence. In A. M. Williams & N. J. Hodges (Eds.), *Skill acquisition in sport: Research, theory and practice* (pp. 231–258). London, UK: Routledge.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*, 158–77.

Watson, J. B. (1920). Is thinking merely the action of language mechanisms? *British Journal of Psychology*, *11*, 87–104.

Wenger, M. J., & Payne, D. G. (1995). On the acquisition of mnemonic skill: Application of skilled memory theory. *Journal of Experimental Psychology: Applied*, *1*, 194–215.

Wilding, J., & Valentine, E. (1997). *Superior memory*. Hove, UK: Psychology Press.

Wundt, W. (1897). *Outlines of psychology* (Translated by C. H. Judd). Leipzig: Wilhelm Engelmann.

Wundt, W. (1907). Über Ausfrageexperimente und über die Methoden zur Psychologie des Denkens [On interrogation experiments and on the methods of the psychology of thinking]. *Philosophische Studien*, *3*, 301–360.

Author Notes

This article was prepared in part with support from the FSCW/Conradi Endowment Fund of Florida State University Foundation. The author wants to thank Robert Hoffman, Katy Nandagopal, and Roy Roring for their valuable comments on an earlier draft of this Chapter.