Specialization Effect and Its Influence on Memory and Problem Solving in Expert Chess Players

Merim Bilalić,[a] Peter McLeod,[a] Fernand Gobet[b]

[a]*Department of Experimental Psychology, Oxford University*
[b]*School of Social Sciences, Brunel University*

Abstract

Expert chess players, specialized in different openings, recalled positions and solved problems within and outside their area of specialization. While their general expertise was at a similar level, players performed better with stimuli from their area of specialization. The effect of specialization on both recall and problem solving was strong enough to override general expertise—players remembering positions and solving problems from their area of specialization performed at around the level of players 1 standard deviation (SD) above them in general skill. Their problem-solving strategy also changed depending on whether the problem was within their area of specialization. When it was, they searched more in depth and less in breadth; with problems outside their area of specialization, the reverse. The knowledge that comes from familiarity with a problem area is more important than general purpose strategies in determining how an expert will tackle it. These results demonstrate the link in experts between problem solving and memory of specific experiences and indicate that the search for context-independent general purpose problem-solving strategies to teach to future experts is unlikely to be successful.

Keywords: Psychology; Memory; Problem solving; Expertise; Reasoning; Pattern recognition; Human experimentation; Problem-solving strategies; Specialization; Thinking; Chess

1. Introduction

How do experts solve problems? Theories of expertise such as chunking theory (Chase & Simon, 1973) and template theory (Gobet & Simon, 1996a) explicitly assume that knowledge of previous problem situations, together with solutions associated with them, plays a major role. This can be seen when chess Grand Masters (GMs) play a number of different

Correspondence should be sent to Merim Bilalić, Tübingen University, Experimental MRI, Department of Neuroradiology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany. E-mail: merim.bilalic@med.uni-tuebingen.de

games simultaneously against weaker players. They do not engage in long deliberations at each board but instead use their extensive knowledge of similar situations to generate moves that are adequate to beat most good club players after looking at the board for a few seconds. However, recent empirical findings and theoretical considerations have questioned the widely held assumption that knowledge is central to expert problem solving (e.g., Chabris & Hearst, 2003; Linhares & Brum, 2007; Schunn & Anderson, 1999).

Another controversial issue is experts' problem-solving strategies. Many researchers on problem solving believe that an understanding of the methods and strategies used by experts is central to the design of successful programs for training future experts (e.g., Anderson, 1993; Newell, 1980; Williams, Papierno, & Makel, 2004). The idea underlying this belief, that there are teachable general thinking skills, applicable across domains, is also held by proponents of the critical thinking movement (e.g., de Bono, 1982; Ennis, 1991, 1996). Given widespread agreement about the importance of discovering experts' problem-solving strategies, it is disappointing to find that research on this topic is uncertain and inconsistent. Do experts rely more on general analytic abilities or on knowledge gained from tackling similar problems? Do they examine many possible solutions (broad search) or do they focus on a single promising solution that they investigate extensively (deep search)? Do they use the same strategies for all problems or does their choice depends on problem characteristics such as difficulty? Are experts' strategies different from those of novices? Do all experts use the same strategies or are there individual differences between the experts themselves? A conclusive answer cannot yet be given to any of these questions despite many decades of research on expertise. This raises doubts over the whole enterprise of trying to discover experts' problem-solving strategies.

In this paper, we will first review inconsistencies in research in which expert problem solving and its link to memory has been studied. We will then propose that the paradigm of ''specialization'' can avoid some of the problems that have led to inconsistent results, and we will present the results of our study with expert chess players using this method. We will show that the effect of memory, that is, familiarity with the sort of problem they are facing, is so strong that the problem-solving performance of expert chess players resembles that of players 1 SD below their skill level when they are taken out of their area of specialization.

1.1. The link between memory and problem solving

Since the seminal study of de Groot (1978) showing that super experts (GMs)[1] have similar patterns of analytical search to ordinary experts (Candidate Masters, CMs) but much more domain-specific knowledge, most expertise researchers have believed that memory is central to successful problem solving. The underlying assumption is that in the course of focused practice, experts encounter and store numerous recurring patterns and successful solutions associated with them. For example, theories that suppose that experts acquire knowledge through chunking mechanisms (chunking theory, Chase & Simon, 1973; template theory, Gobet & Simon, 1996a) propose a direct link between memory, as captured by recall tasks, and problem-solving ability. The knowledge base of acquired chunks and more complex templates, which can be seen as prototypical problems/positions, steadily grows

and becomes increasingly differentiated through practice as do the possible actions connected with them. Chunks and templates—which we will call ''knowledge structures''—become the link between pattern recognition and higher-level conceptual knowledge. Once pattern recognition processes have identified a problem as familiar, information about the problem, including potential ways of dealing with it, is drawn from long-term memory.

The idea that knowledge structures play a key role in the development of expertise has led to the development of computational models. For example, the CHREST (Chunk Hierarchy and REtrieval STructures) model, a partial implementation of template theory, has been applied to chess (e.g., de Groot & Gobet, 1996; Waters & Gobet, 2008) and to awalé, an African board game (Gobet, in press). The program learns by (a) acquiring perceptual chunks, which relate to patterns of pieces on the board; (b) learning possible moves and sequences of moves; and (c) associating moves with perceptual chunks. CHREST has simulated a number of phenomena about memory and problem solving in these two games and has also simulated the differences between the eye movements of weak players and masters in chess.

The evidence for the view that memory plays a central role in expert problem solving is abundant. First, there are clear-cut differences in the amount and organization of knowledge in experts and novices (Chase & Simon, 1973; Chi, Glaser, & Farr, 1988; de Groot, 1978). Second, there are negligible differences in search strategies of super and ordinary experts (Charness, 1989; Gobet, 1998b; de Groot, 1978), which points to the importance of pattern-recognition processes as an explanation of super experts' superior performance (Burns, 2004; Charness, 1989; Gobet & Simon, 1996b). Third, in simultaneous play, where an expert plays a number of weaker opponents at the same time and thus has much less time to think about each move than it is usually the case, the best players still perform formidably well. For example, the former World champion Gary Kasparov beat all but one member of the Swiss National Team in a simultaneous exhibition (Gobet & Simon, 1996b). Similarly, there are indications that, at higher skill levels, pattern recognition plays a more important role than analytical processes, such as search, while the analytical processes are more important for weaker players. Burns (2004) showed that, in rapid games, where the thinking time is severely limited (typically to a few seconds per move) and thus lengthy search processes are prevented, the differences among strong players are roughly the same as in normal games where they have plenty of time to search (typically an average of 3 min per move). On the other hand, the performance in rapid games among weaker players does not correlate highly with the performance in normal games.

Although the prevailing view is that knowledge structures acquired through extensive practice lead to superior performance, there are alternative views. The role of templates and chunks in the problem-solving process has recently been deemphasized in Linhares' theoretical and empirical work (Linhares, 2005; Linhares & Brum, 2007). According to Linhares, strong players form ''abstract roles'' based on the deep meaning of the board constellations and not on the surface appearance as in the template and chunking theories. The same abstract concepts can be found in different positions that do not necessarily share the templates and chunks in the classical sense. Holding (1985, 1992) claimed that the main factor

of chess skill is forward search, analytical reasoning skill, and not pattern recognition. As Holding (1985, pp. 249–250) put it: ''There is no doubt that experienced players possess extremely rich and highly organized chess memories, but the most useful attributes of these memories seem to be more general than specific and, if specific, not necessarily concerned with chunked patterns.'' According to Holding, stronger players not only search more widely and more deeply than weaker players but they also make use of evaluation to differentiate between different paths/solutions. Holding's SEEK (SEarch, Evaluation, and Knowledge) theory includes a knowledge component that guides forward search, but its importance is overcome by search and evaluation.

There is also evidence that superior performance can be achieved without extensive practice. Ericsson and colleagues (Ericsson, 1985; Ericsson & Chase, 1982; Ericsson & Harris, 1990; Ericsson & Oliver, 1989) demonstrated that, with a mere 50 h of practice, people could reach the digit-span level of professional memory experts with over 20 years of experience (Ericsson & Chase, 1982) or be able to recall unfamiliar chess stimuli as well as experienced experts with thousands of hours of chess practice (Ericsson & Harris, 1990; Ericsson & Oliver, 1989; see also Ericsson & Lehmann, 1996, for a review of other ''dissociations'' between memory and problem-solving performance). As Ericsson and Kintsch (2000) put it, ''If expert memory performance can be attained in a fraction of the number of years necessary to acquire expert chess-playing skill, then this raises doubts about the necessity of *a tight connection between expert performance and experts' superior memory for representative stimuli*'' (p. 578; emphasis added).

1.2. Problem-solving strategies in chess

When confronted with a novel problem, solvers have to decide, consciously or unconsciously, whether they will examine a small number of alternatives in depth, or whether they will consider many different solutions and investigate them all to a lesser extent (depth search vs. breadth search). The common wisdom is that search in depth is faster, more efficient, and less demanding for memory than breadth search where it is necessary to set goals and subgoals and keep track of them throughout problem solving (Larkin, McDermott, Simon, & Simon, 1980; Newell & Simon, 1972; Patel & Groen, 1986). On the other hand, depth search may be a risky strategy because the search is executed without first checking whether it is relevant to the main goal (Hunt, 1991). It may be efficient for experts (who will be familiar with solutions to similar problems and so likely to choose an effective method) but not so good for novices who are less familiar with the domain.

de Groot's (1978) research showed that all chess players first became familiar with the problem, identified goals, and related them to their knowledge. This process enabled them to generate specific methods of tackling the problem, which were, in turn, investigated employing search strategies (see Chase & Simon, 1973, and Saariluoma, 1995, for detailed and elaborated mechanisms of the whole process). The most surprising result was that there were no significant differences in the macrostructure of the problem-solving strategies used, dependent on the level of expertise. Super experts (some of the strongest GMs at the time) and ordinary experts (CMs—see Note 1) did not have different preferences when it came to

depth of search and breadth of search. Both groups investigated a similar number of positions and solutions (measures of breadth of search), had a similar maximal depth of search and searched on average to a similar depth (measures of depth of search), and reinvestigated solutions to the same extent. Despite these similarities, super experts did, however, find better solutions than ordinary experts (for a review, see Bilalić, McLeod, & Gobet, 2008c).

The surprising finding of no differences in the macrostructure of search between skill levels may have been due to the small number of participants and their limited skill span (<3 SDs of range in skill from the best players [super experts] to the weakest [ordinary experts]) (see Holding, 1985). Subsequent studies, using the same position and procedure as de Groot, but a wider skill span showed that there are differences between experts and nonexperts in the macrostructure of search. Gobet (1998b) showed that Masters (M) (4 SDs above the mean of all players) do search more deeply on the average than Class B players (1 SD above the mean), but there were no differences between Ms, CMs (3 SDs above the mean), and Class A players (2 SDs above the mean). Similarly, the CMs (about 3 SDs above the mean) in the study by Gruber (1991) searched more deeply than novices, although there were no differences in the breadth of search (e.g., number of candidate moves considered).

Other studies, using different positions and time limits, indicate that there are indeed differences between strong players (3 SDs above the mean) and weak players (a couple of SDs below the mean) in the structure of search. Mean depth increased by 1.5 ply (a ply is one move by one player, sometimes called a half-move) with every SD used in the study by Charness (1981). Players with a rating of 1,300 (1 SD below the mean) searched on average 3.6 ply in comparison with 9.1 ply for the most skilled players in the study (3 SDs above the mean). Based on these results, Charness (1981) suggested that the depth of search increases with increase in skill until about expert level (2,000 Elo, or 2.5 SDs above the mean), after which it remains uniform. In the only longitudinal study on problem solving in chess, Charness (1989) found that a participant from his earlier study (Charness, 1981) did not show an increase in depth of search despite the fact that he had improved from an average player to an International Master (IM) 9 years later. The player in question did, however, display a more compact search pattern: He spent less time on the positions and investigated fewer candidate moves.

These results led to the conclusion that problem-solving strategies are important for average players but are less relevant for highly skilled players (Charness, 1989; Gobet, 1998b; de Groot, 1978). For example, template theory (Gobet & Simon, 1996a; implemented into a simulation program for search in chess, SEARCH, Gobet, 1997) predicts that the average depth of search should follow a power function—at lower skill levels the increase in the depth of search should rapidly follow increase in skill, but as skill level increases, the increase in the depth of search should become less and less. Consequently, one of the corner-stones of theories of expertise is that recognition processes based on knowledge are more important than analytical processes, such as search, for experts' performance (see Gobet, 1998a, for a review).

Although these results suggest that there are differences in problem-solving strategies between experts and novices, but that among experts those differences are largely overshadowed by knowledge, there are several studies that have come to a different conclusion.

Saariluoma (1992, Experiment 3) found that Ms searched more broadly and more deeply than weak players in an endgame position, but in tactical positions, where a winning combination is usually available, IM and GMs' search was narrower than that of Ms and Class A players (Saariluoma, 1990, Exp. 5). Chabris and Hearst (2003) established that preventing search processes (as in rapid games) had a deteriorating effect on the performance of the very best chess players, while van Harreveld, Wagenmakers, and van der Maas (2007) could not replicate Burns' (2004) finding among elite chess players—search processes were as important for the very best players as they were for their weaker colleagues.

1.3. Methodological problems of previous studies

In short, research on the problem-solving strategies that expert chess players use has produced contradictory results. Similar contradictory findings have also been found in physics (Clarke & Lamberts, 1997; Larkin et al., 1980), design (Ball, Evans, Dennis, & Ormerod, 1997; Jeffries, Turner, Polson, & Atwood, 1981), and medicine (Elstein, Shulman, & Sprafka, 1978; Kulatunga-Moruzi, Brooks, & Norman, 2001; Patel, Groen, & Arocha, 1990). If the question of the influence of knowledge and memory on problem-solving problem solving (strategies) is at the heart of the investigation into the nature of expertise, and is also required to provide the best training of future experts, this confusion is highly unsatisfactory. If expert performance is not dependent on superior memory and knowledge, then more emphasis should be put on techniques that train analytical skills than on the acquisition of knowledge through practice.

It is possible that the confusing results are a consequence of the methods employed. Often different time constrains, difficulty of problems, and different scoring systems were used in different studies. We also believe that the paradigm of comparing experts and novices used in most studies is inherently plagued with problems that prevent us from drawing valid conclusions. First, it is not agreed who are experts and who are novices. Secondly, besides the difference in expertise, experts and novices usually differ on other characteristics such as age, education, and, in particular, motivation for the task. Finally and most importantly, it is difficult to find suitable problems because of the difference between experts and novices (Reimann & Chi, 1989). It is likely that an appropriate problem for experts would be too difficult for novices, whereas the appropriate one for novices would be too easy for experts.

1.4. The specialization paradigm

The specialization paradigm offers a possible way of avoiding the confounds in the expert–novice paradigm. Instead of comparing experts with novices, two groups of experts with different fields of specialization are compared. The two groups will, therefore, have similar experience and general skill level but different knowledge bases, allowing the effects of familiarity with the problem type (memory) and general experience (skill) to be teased apart. The specialization paradigm has been previously applied in medicine (Joseph & Patel, 1990), political science (Voss, Tyler, & Yengo, 1983), and experimental design domain (Schraagen, 1993; Schunn & Anderson, 1999). For example, Schraagen (1993) showed that

domain experts (with 10 or more years of experience in designing experiments in the area of the problem) and design experts (with 10 or more years of experience with designing experiments in psychology but outside the area of the problem) display similar problem-solving strategies that are in contrast with the way undergraduate and graduates students tackle the problem. Similarly, in the study by Schunn and Anderson (1999), domain experts and task experts used domain-general strategies to the same extent but domain experts displayed a greater use of domain-specific strategies. Undergraduates, on the other hand, lacked knowledge of both domain-general and domain-specific strategies.

These results show that even when the necessary domain knowledge is lacking, experts can revert to general strategies to deal with the problem. As these general strategies are not found among novices, Schunn and Anderson (1999) claimed that this result contradicts the main assumption in theories of expertise ''that domain expertise consists primarily of a large quantity of domain-specific facts, skills, and schemata acquired only through thousands of hours of practice'' (p. 366). The authors further concluded that ''expertise … may also consist of many domain-general skills.'' Similarly, Schraagen (1993) states that ''experts have flexibility that goes beyond mere domain specific knowledge. When this knowledge is lacking, experts can still maintain a more structured approach than novices by making use of more abstract knowledge and strategies'' (p. 305).

There are, however, methodological shortcomings of the studies involving differently specialized experts, which cast doubt on the conclusions. In the studies by Schraagen (1993), Schunn and Anderson (1999), and Voss et al. (1983), as well as the studies of medicine subexperts (Joseph & Patel, 1990), usually only one problem was presented. The problem is necessarily from the area of one group of experts but outside the area of the other group of experts. To control for differences between experts themselves, it is necessary to give two kinds of problems—one from the area of each group. In doing so, it is possible to check that the same pattern is observed with both groups of experts. Similarly, none of the studies used neutral problems outside the areas of specialization of all participants. Neutral problems act as a control for different skill levels within specialization groups and provide further insight into the generality of the problem-solving strategies observed when experts were in their area of specialization. Finally, in all studies, it was never clear how good the experts were in comparison with the novices and subexperts (i.e., experts outside their area of specialization).

1.5. Overview of the study

In our study, we wanted to overcome the methodological problems identified in previous research, which compared experts to novices (by using the specialization paradigm), and in previous use of the specialization paradigm (by using problems in both areas of specialization and neutral problems). The specialization paradigm can be used with chess because it is a complex domain where experts have their own subareas of specialization and it offers a reliable and objective measure of skill (the Elo scale) to balance the levels of expertise in the different specialization groups. Two types of players participated in our study. The first group specialized in one opening (the French defense), while the second group specialized

in another (the Sicilian defense). (Different openings lead to different sorts of position so players tend to specialize in certain openings and ignore those they know they will not play. The decision to follow the French or Sicilian defense is a decision made by the second player [Black] in response to an opening move of pawn to e4 by White. If Black chooses to reply by moving a pawn to e6, the game becomes a French; if the choice is to move a pawn to c5 it becomes a Sicilian.) Both groups were similar in general skill level. The same groups of players first recalled positions and then solved problems within their area of specialization, outside their area of specialization, and with neutral problems. The ''neutral'' problems came from middle-game positions, so should not be influenced by opening specialization but reflect more general memory and problem-solving abilities.

Theories in which expertise is based on chunking mechanisms (e.g., Chase & Simon, 1973; Gobet & Simon, 1996a) predict that players with different specialization will possess knowledge bases that will have dissimilar elements, as the players have been exposed to different stimuli during their chess career. As a consequence of differently specialized knowledge, players should remember positions and solve problems within their area of specialization better but have approximately equal success with the neutral problems. If general expertise and analytical abilities are more important (Holding, 1985, 1992; Patel & Groen, 1991), or if players are able to form similar abstract concepts from different positions as Linhares (2005; Linhares & Brum, 2007) suggests, then different problems may produce differences in experts' ability to recall the position, but there should not be marked differences in the quality of chosen solution in problem solving.

Problem search can be characterized by depth and breadth of search. The depth of search measures indicates how far/deep the solver investigates a particular solution, while the breadth of search specifies how many possible solutions the problem solver considers. The main measures of depth are the maximum reached for any solution and the average across the solutions tried. Breadth of search is predominantly defined through the number of different solutions tried out. If there is a uniform strategy used by all experts, it should be reflected in similar measures of depth and breadth of search on all problems, whether they are within their area of specialization or outside it. On the other hand, if familiarity with the problem influences the strategies in use, one can expect different behavior depending on the problem. Experts should try out more solutions on problems they are less familiar with. One of the consequences of considering more solutions is that some will inevitably turn out to be unproductive and will be abandoned after only a short investigation. Therefore, problems outside the area of specialization should elicit more extensive search but the search will, on average, be shallower. Problems within the area of specialization will not force experts to search extensively because they are already familiar with common plans. They will therefore concentrate on a few possibilities, which will be investigated in depth. Consequently, if familiarity with the problem type influences search style, it is expected that problems within the area of specialization will elicit greater depth but less breadth, whereas the pattern will be opposite for the problems outside the area of specialization.

By using experts of different skill levels (CMs, Ms, and IM&GMs), it will be possible to investigate whether specialization can override the influence of expertise. That is, weaker players in their area of specialization may outperform stronger players on the same

problems, who are outside their area. Hence, including experts of different skill levels makes it possible to disentangle the relation between knowledge and search behavior as well as to clarify whether there are uniform problem-solving strategies that most experts employ. Finally, including Neutral problems, taken from the middle game, in which the influences of opening specialization should be less marked, will enable us to see whether differences found within and outside the area of specialization continue when the influences of specialization are no longer there.

2. Method

2.1. Participants

Players who specialize in playing either the French or Sicilian defense participated in the experiment. The French and Sicilian defense were chosen because they are among the most popular openings that enabled us to recruit a decent number of experts. There were three skill levels within each group: CM, M, and IM&GM. Players were recruited either during the Bosnian team championship in 2003 and 2004, or through personal contacts of the first author. Table 1 shows the average ratings and age within skill levels and specialization groups. There were no significant differences in rating and age between the two groups of players (nor an interaction between skill and age of players for rating).

2.2. Stimuli

Four types of positions were used as problems: Sicilian, French, Neutral, and Random. The first three types contained four different examples, and the last had two. The Sicilian and French positions were taken from a specific line from the opening (the Najdorf for the Sicilian, the Winawer for the French). The French positions had on average 28.5 pieces

Table 1
Mean and standard deviation, *M* (*SD*), of players' Elo rating and age

Player Type/Skill Level	Rating	Age	*n*
French			
Candidate Master	2,132 (57)	22 (2)	4
Master	2,299 (18)	35 (13)	4
IM&GM	2,452 (35)	37 (7)	4
Total	2,294 (141)	31 (10)	12
Sicilian			
Candidate Master	2,141 (79)	28 (12)	4
Master	2,305 (45)	35 (16)	4
IM&GM	2,520 (101)	30 (11)	4
Total	2,322 (177)	31 (12)	12

Note: IM&GM, International and Grand Master.

(*SD* = 1.8), while the Sicilian positions had 29 (*SD* = 1.9). Neutral positions were taken from middle-game positions played by lesser-known masters. The Neutral positions originated from openings other than the French or Sicilian. All four Neutral positions had 28 pieces. The two random positions were generated so that any kind of piece could appear on any square, with no restriction on the distribution of pieces (Gobet & Waters, 2003; Vicente & Wang, 1998). Both random positions had 26 pieces. Fig. 1 shows examples of positions used with the best solutions. The complete set of positions can be obtained from the first author.

2.3. Familiarity

To identify players who play the Sicilian or French defense, but not both, we employed a familiarity questionnaire with 16 positions. Some positions were from the French or Sicilian defense, but they were mixed up with other unrelated positions and drawn randomly from the pool of chess openings, to avoid suspicion about the purpose of the questionnaire. The

French position (1...Rg6)   Sicilian position (1...Ne8)

Neutral position (1...c5)   Random position

Fig. 1. Examples of the positions used in the memory and problem-solving studies (clockwise—French, Sicilian, Random, and Neutral position). In the problem-solving studies, it is Black to move. The best move is shown in brackets.

full questionnaire can be obtained at http://www.cogsci.rpi.edu/CSJarchive/Supplemental/ index.html. Participants were asked how frequently they played the particular opening featured in the position using a scale anchored at 1 (Never) and 6 (Always). The French and Sicilian positions used were from the same type of opening as those in the experiment, but they were also markedly different from the actual positions used. The openings in question were broad enough that even the best experts cannot know all the lines and sublines. Consequently, there is no guarantee that even the players who are specialized in some of the lines of broad openings such as the French and Sicilian defenses will be able to rely on their previous memories. Table 2 presents the answers to the questions about the playing frequency (described later in the text as ''familiarity'') of Sicilian and French players for the French and Sicilian positions. Most players specialized in one opening indicated that they hardly (1 or 2 on the scale) ever played the other opening except a few strong Sicilian players who occasionally played the French opening too (3 on the scale). Players who scored at least 4 (often) on one opening and <3 (rarely) on the other participated in the experiments.

A mixed ANOVA with player type, position type, and skill level as fixed between factors and players and positions as nested random repeated factors on the frequency of playing the particular lines featured in the positions was performed. The interaction player type × position type was highly significant ($F[1, 164] = 451.91$, MSE = 0.88, $p < .01$, $\eta_p^2 = 0.73$) confirming the obvious result that players played the positions within their area of specialization more often than outside. We used an independent measure of playing frequency to validate the subjective familiarity ratings. With the help of the ChessBase database (ChessBase GmbH, Hamburg, Germany), which contains over 2 million games, we found all games in which the players had the Black pieces and faced 1.e4 as the first move (to which either the Sicilian or French defenses would be possible replies). We then calculated the percentage of time each player used the French or Sicilian defense as his or her response.

Table 2

Mean and standard deviation, *M* (*SD*), of playing frequency of the opening lines featured in the stimuli positions

Player Type/Skill	Position Type	
	French	Sicilian
French		
Candidate Master	3.9 (1.8)	1.2 (0.4)
Master	4.4 (1.4)	1.6 (1)
IM&GM	4 (1.5)	1.1 (0.3)
Total	4.1 (1.5)	1.3 (0.7)
Sicilian		
Candidate Master	1 (0)	4.3 (1.4)
Master	1 (0)	5.3 (1.3)
IM&GM	2.2 (1.4)	4.9 (0.6)
Total	1.4 (1)	4.8 (1.2)

Note: 1 = never; 6 = always.
IM&GM, International and Grand Master.

Unsurprisingly, the Sicilian players predominately used the Sicilian defense (81%), while they chose the French defense only 6% of the time. The French players had the opposite preference—84% French and only 3% Sicilian.

2.4. Design and procedure

2.4.1. Memory

Positions were presented on an 8″-screen portable Apple MacIntosh computer using specialized software for presenting chess stimuli and recording responses (for more details about the software, see Gobet & Simon, 1998). Participants first familiarized themselves with the computer display and were shown how to select and place pieces on the board. They then received two middle-game positions for practice. Each position was presented for 5 s after which the board went blank and the player tried to reconstruct the position from memory on an empty board on the screen. After the practice, 12 stimulus positions (French, Sicilian, and Neutral) were presented, each shown for 5 s with no time limit for recalling a position. The presentation order of the game positions (French, Sicilian, and Neutral) was random for each participant. After the game positions, two Random positions for practice, followed by another two Random positions, were presented.

2.4.2. Problem solving

After the recall task and a short break, the participants were given the problem-solving task. Participants were individually tested using the think-aloud procedure (Ericsson & Simon, 1993). They read the instructions, which stated that they should look for the best move in the positions and that they had 10 min to do so. The two French, two Sicilian, and two Neutral positions that players already had recalled were presented in a random order. All positions were shown on a 15″-screen laptop computer using ChessBase, a standard chess program most players are familiar with. To emulate the natural tournament situation, the participants were not allowed to move the pieces. They were tested individually in a quiet room and the whole problem-solving session was tape recorded. The participants took about an hour to find solutions for all six problems.

2.5. Analysis

2.5.1. Move quality

The move quality in the problem-solving task was established using Fritz 8, a strong chess program.[2] Fritz 8 gives evaluations of moves in pawn units (e.g., +0.5 means that White has an advantage of half a pawn). Given that one position could be better for Black from the start while another could favor White (e.g., the best move in one position could be −1.19, that is, Black is better by 1.19 pawns, while in the other the best move would produce an assessment of +0.06 where White is better by 0.06 of a pawn, that is, even with the best move selected by Black, White's position is still superior), we measured the absolute difference in pawn units from the best move in the position. Hence, an assessment of 0.2 means that the selected move was inferior by 0.2 of a pawn to the best move for Black in that position.

2.5.2. Protocol parameters

The verbal protocols were used to construct problem behavior graphs (PBGs; Newell & Simon, 1972) for every player. Besides the exact time and the final solution, it was possible to extract several other parameters from PBGs. The player starts with mentioning a move, a possible solution which we will call a candidate move. The player then investigates the path that is opened with the move. This investigation of a path, which starts with a candidate move and follows by other moves in a sequence, is called an episode. During an episode, the player can investigate different subpaths within the same episode. A move in the episode can have two possible replies, which lead to two different branches of the same episode. The episode is concluded when the player comes back to the initial position. The player can then investigate another solution, which would count toward the number of candidate moves, or can reinvestigate the previous candidate move. It is also possible to calculate the total number of moves mentioned during the search process as well as the speed of problem processing, which represents the number of moves investigated per minute. Two parameters of depth of search can also be obtained from the protocols. Average depth of search, or mean depth, shows how many half-moves (ply) on average were considered during the search. The other depth of search parameter is the maximal depth of search, which represents the greatest depth reached during the search in half-moves.

Although it is customary in research on chess problem solving to talk about the depth of search, breadth of search is rarely mentioned. As we wanted to look how these two problem-solving strategies are influenced by the context of familiarity, we conducted factor analysis on all three different types of problems. These analyses, presented in the Appendix, identified two groups of variables. One had mean and maximum depth parameters together, while the other group included the number of candidate moves and episodes. The other parameters were not sufficiently stable over different types of positions to be included in the depth and breadth of search categories. We will briefly summarize the analyses of the other parameters in the main text but will not present the detailed analyses.

2.5.3. Statistical analysis

We transformed the percentage of correctly recalled pieces using the arcsin function to obtain approximately normal distributions. Given that we were interested in the specialized stimuli (French and Sicilian positions) and used the Neutral and Random stimuli as controls, we analyzed the specialized stimuli together using a mixed ANOVA with player type (Sicilian and French), position type (Sicilian and French), and skill level (IM&GM, M, and CM) as fixed between factors and players and positions as nested random repeated factors. There were no differences between performing ANOVAs on the two specialized position types alone on the one hand, and together with the Neutral stimuli on the other, except that the Neutral stimuli were harder to remember than the specialized stimuli. This is not surprising because the Neutral positions were less structured, being taken from middle-game positions, while the French and Sicilian positions were late opening or early middle-game positions that resulted in more familiar structures. The control stimuli were thus analyzed separately using ANOVAs with player type and skill level as between factors and positions as repeated measures.

Finally, the effect size for ANOVAS was estimated using η_p^2, which is the proportion of the cumulative variance of effect and error that is attributable to the effect. For t-tests, we used Cohen's d, which represents the difference divided by the pooled SD of both means.

3. Results and discussion

3.1. Memory

French and Sicilian players were equally successful (as measured by arcsin-transformed percentage of successfully recalled pieces) when performance was pooled across the specialized positions and there were no differences in how well French and Sicilian positions were recalled (Table 3) Unsurprisingly, more skilled players recalled the positions better than less skilled players ($F[2, 18] = 19.88$, MSE = 259.67, $p < .01$, $\eta_p^2 = 0.69$). The skill effect was also apparent with the Neutral stimuli—more skillful players outperformed their less skillful colleagues ($F[2, 18] = 6.94$, MSE = 201.01, $p < .01$, $\eta_p^2 = 0.44$). There were no differences in recall of the Random positions between the groups. Given that the recall of random position is dependent on general memory abilities (Chase & Simon, 1973; Gobet & Waters, 2003), this suggests that there was no difference in general memory abilities between the two groups of specialized players. Similarly, skill had no impact on the recall of random positions nor there was an interaction between skill and player type.

The crucial result is that the players were better at recalling positions within than outside their opening of specialization. French players were better at recalling the French positions; Sicilian players were better at recalling the Sicilian positions. The interaction between player and position types was significant ($F[1, 156] = 46.96$, MSE = 61.82, $p < .01$,

Table 3
Transformed (arcsin) percentage and standard deviation, M (SD), of correctly recalled pieces in French, Sicilian, Neutral, and Random positions as a function of the group and skill level of players

Player Type/Skill Level	Position Type			
	French	Sicilian	Neutral	Random
French				
Candidate Master	71 (8)	63 (11)	51 (9)	26 (6)
Master	75 (11)	68 (8)	57 (9)	28 (10)
IM&GM	81 (8)	79 (8)	60 (7)	28 (3)
Total	75 (10)	70 (11)	56 (9)	27 (7)
Sicilian				
Candidate Master	57 (13)	67 (9)	51 (7)	28 (11)
Master	67 (9)	80 (9)	60 (9)	30 (9)
IM&GM	81 (10)	89 (2)	68 (13)	28 (13)
Total	69 (14)	79 (12)	59 (12)	29 (10)

Note: IM&GM, International and Grand Master.

$\eta_p^2 = 0.23$). With the French and Sicilian positions, familiarity managed to override skill in that less skillful players presented with a position from within their area of specialization performed as well as more skillful colleagues when that problem was outside their area of specialization. The performance of Sicilian CMs and Ms on Sicilian positions (67% and 80%) was comparable to that of French Ms and IM&GMs, respectively (68% and 79%), that is, to players 1 SD above them in skill (see Note 1). Similarly, French CMs and Ms (71% and 75%) recalled French positions on average as well as Sicilian Ms and IM&GMs (67% and 81%).

The extent of the specialization effect on chess memory seems to be around 1 SD—chess players recalling positions within their opening of specialization performed at a similar level to players who were 1 SD above them in skill but were dealing with positions outside their opening of specialization.

3.2. Problem solving

There were no significant differences due to player or position type (see Table 4). The neutral middle-game positions were solved at the same level by both groups confirming that both groups were of a similar skill level. As would be expected, more skilled players chose better moves (F[2, 18] = 7.35, MSE = 0.34, $p < .01$, $\eta_p^2 = 0.45$). More skilled players also solved the Neutral position better than their less skilled colleagues (F[2, 18] = 7.68, MSE = 0.13, $p < .01$, $\eta_p^2 = 0.46$).

As in the memory experiment, the crucial result is the interaction between player and position types. Players who were in their opening specialization produced better solutions than those who were outside it (F[1, 64] = 13.87, MSE = 0.30, $p < .01$, $\eta_p^2 = 0.18$).[3] The extent of the specialization effect was similar to that observed in the memory task. On French problems French CMs ($M = 1.05$) and Ms ($M = 0.31$) performed slightly better than Sicilian Ms ($M = 1.19$) and IM&GMs ($M = 0.47$), respectively. That is, the French players performed at the level of Sicilian players 1 SD above them in skill. With the Sicilian problems, the effect was even more marked. Sicilian CMs were better at solving Sicilian positions ($M = 0.23$) than the French IM&GMs ($M = 0.32$) who were 2 SDs above them in skill.

3.2.1. Problem-solving strategies on stimuli within and outside the area of specialization

The analysis of the protocols showed that the problem-solving strategy used depended on the problem type. When confronted with the problems within their specialization, players' search pattern inclined more toward depth and less toward breadth. The same players employed the opposite search pattern on the positions outside the opening of their specialization—they examined more candidate moves and generated more episodes but exhibited shallower depth of search (Table 5). This resulted in a significant interaction between player and position types for mean depth (F[1, 64] = 9.73, MSE = 1.68, $p = .003$, $\eta_p^2 = 0.13$), max depth (F[1, 64] = 5.06, MSE = 6.46, $p = .028$, $\eta_p^2 = 0.07$), candidate moves (F[1, 64] = 7.21, MSE = 3.19, $p = .009$, $\eta_p^2 = 0.10$), and episodes (F[1, 64] = 4.54, MSE = 15.43, $p = .037$, $\eta_p^2 = 0.07$). The analyses of other protocol parameters showed that

Table 4

Mean and standard deviation, *M* (*SD*), of solution quality on French, Sicilian, and Neutral positions as a function of the group and skill level of players

Player Type/Skill Level	Position Type		
	French	Sicilian	Neutral
French			
Candidate Master	1.05 (0.85)	0.66 (0.48)	1.18 (0.39)
Master	0.31 (0.43)	0.46 (0.61)	1.04 (0.66)
IM&GM	0.11 (0.32)	0.32 (0.22)	0.89 (0.61)
Total	0.49 (0.69)	0.48 (0.47)	1.04 (0.55)
Sicilian			
Candidate Master	1.23 (0.85)	0.23 (0.24)	1.38 (0.19)
Master	1.19 (0.87)	0.14 (0.20)	0.95 (0.62)
IM&GM	0.47 (0.71)	0.02 (0.07)	0.68 (0.74)
Total	0.97 (0.85)	0.13 (0.20)	1.01 (0.62)

Note: The numbers indicate the deviation from the best solution on a scale where 1 is the value of a pawn. Smaller values denote better solutions with 0 being the best solution.

    IM&GM, International and Grand Master.

players also spent less time and reinvestigated the candidate moves less often on the positions within the opening of specialization. These differences indicate that the problems within the opening of specialization were easier to tackle than the problems from outside the opening of specialization. The players had a better idea of likely good moves on the problems within their area of specialization and hence looked at fewer candidate moves, and were able to investigate the moves they did consider to greater depth than on the problems outside their area of specialization. Thus, they were likely to find better solutions.

Skill level was significant for depth (mean depth, $F[2, 18] = 4.79$, MSE $= 1.52$, $p = .021$, $\eta_p^2 = 0.35$; maximal depth, $F[2, 18] = 5.50$, MSE $= 9.93$, $p = .014$, $\eta_p^2 = 0.38$) in that IM&GMs searched to significantly greater depths than Ms. There were no significant differences between other skill levels (Table 5). The same pattern where IM&GM had higher values than Ms was observed for the breadth of search but it just failed to reach significance, probably due to insufficient power (candidates—$F[2, 18] = 3.26$, MSE $= 2.03$, $p = .062$, $\eta_p^2 = 0.27$; episodes—$F[2, 18] = 3.47$, MSE $= 29.83$, $p = .053$, $\eta_p^2 = 0.28$). A possible reason for this pattern of results is different familiarity with the stimuli. Masters in both groups showed the highest familiarity with the positions within their area of specialization, which might have influenced the amount of effort necessary to investing in problem solving (Table 5).

3.2.2. Problem-solving strategies on neutral stimuli (middle-game positions)

Given that the Neutral problems, unlike the problems from the French and Sicilian openings, yielded no overall differences in protocol parameters between the groups, we pooled the protocol parameters across the groups. Neutral problem 1 was relatively straightforward. It included a clear motif but required deep search for the correct eval-

Table 5

Average values for depth (mean and maximal) and breadth (candidates and episodes) of search of French and Sicilian players on French and Sicilian positions depending on their skill level

	Position Type							
	French				Sicilian			
	Depth		Breadth		Depth		Breadth	
Player Type/Skill Level	*M*	Max	Can	Ep	*M*	Max	Can	Ep
French								
Candidate Master	4.6	8.1	2.6	6.1	3.8	6.3	2.8	6.3
Master	3.8	6.1	1.8	3.6	3.1	6.5	3.6	7.5
IM&GM	4.6	8.1	2.9	5.8	4.3	7.4	2.5	7.6
Total	4.3	7.5	2.4	5.2	3.7	6.7	3	7.1
Sicilian								
Candidate Master	3.2	4.9	3.3	6.1	5.2	7.6	1.9	4.3
Master	3.5	5.5	2.8	5	3.8	4.6	1.8	2.5
IM&GM	4.1	7.4	5	9.6	5	10.3	3.1	9.6
Total	3.6	5.9	3.7	6.9	4.7	7.5	2.3	5.5

Note: IM&GM, International and Grand Master; *M*, mean depth of search; Max, maximal depth of search; Can, candidate move; Ep, episode.

uation of the solution. In contrast, Neutral problem 2 was more difficult, an atypical problem with no clear motif or way of proceeding.[4] One could say that the first problem was within the specialization of all experts, while the second was outside everyone's area of specialization.

Fig. 2 shows the problem-solving strategies on problem 1 for different skill levels. The first problem not only yielded clear differences between the skill levels in the solution quality (GM [0.24] solved problems more successfully than *M* [0.56] who solved it better than CM [1.04]—$F[2, 18] = 6.37$, MSE = 0.21, $p = .008$, $\eta_p^2 = 0.42$) but also both in mean depth ($F[2, 18] = 3.60$, MSE = 4.69, $p = .048$, $\eta_p^2 = 0.29$) and maximal depth ($F[2, 18] = 2.92$, MSE = 11.18, $p = .080$, $\eta_p^2 = 0.24$). While there were no significant differences among skill levels in the breadth of search, all other relevant protocol statistics (e.g., moves per minute) were in linear association with skill.

The second neutral problem was more difficult than the first and the quality of solutions was lower. Now there were no significant differences in mean depth, maximum depth, number of candidate moves, or number of episodes between skill levels (Fig. 3). Although there were indications that more skilled players solved the problem better (1.34, 1.43, and 1.52 for GM&IM, M, and CM, respectively), searched more extensively, and processed the problem faster, none of these differences was statistically significant.

The difference between the two Neutral problems was striking and underlines the importance of familiarity. The players solved the first problem better ($t[23] = 7.7$, $p < .001$, $d = 2.08$), searched deeper on average ($t[23] = 3.9$, $p = .001$, $d = 1.04$), and reached higher maximal depth ($t[23] = 2.4$, $p = .026$, $d = 0.53$). In the second problem, however, players tried more solutions ($t[23] = 4.9$, $p < .001$, $d = 1.32$) and generated more episodes
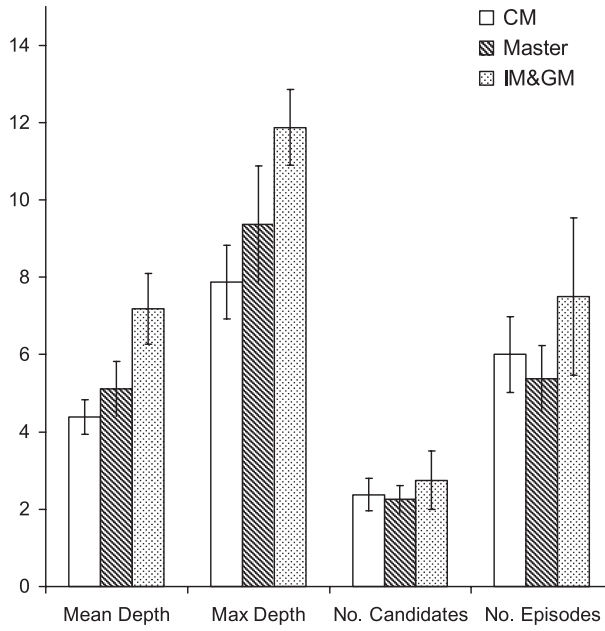
Fig. 2. The parameters of depth and breadth of search for players of different skill on Neutral problem 1. Error bars present standard error of the mean.
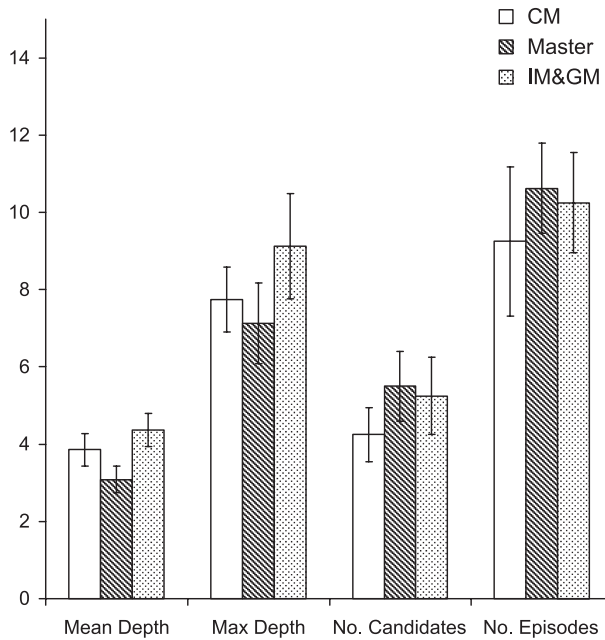

Fig. 3. The parameters of depth and breadth of search for players of different skill on Neutral problem 2. Error bars present standard error of the mean.

(t[23] $= 4.1, p < .001, d = 0.98$). It thus seems that when there is a clear line of analysis as in the first Neutral problem, the more skilled players are, the more likely they are to find it and research it in greater depths (and consequently get a better result) (Fig. 2). On the other hand, when there is no clear motif of play as in the second Neutral problem, differences associated with skill level are less clear in problem-solving strategies (as well as in solution quality) (Fig. 3). Suddenly, with the absence of familiarity, the problem-solving strategies of super experts (GMs) and ordinary experts (CMs) resemble each other just as in de Groot's study.

4. General discussion

Focused experience leads to acquisition of knowledge about a domain, its structure, common problems, and ways of dealing with those problems. Unsurprisingly, people with a vast domain experience (experts) are able to solve problems and remember stimuli from the domain better than people with a limited experience (novices). In this study, we used the specialization paradigm showing that, even among experts, with the same level of general expertise, there are differences that are connected with specific focused experience. Expert chess players both remembered and solved problems arising from their area of opening specialization better than problems outside their specialization. We were also able to quantify the specialization effect—players remembered and solved the problem stimuli within their specialization roughly at the level of players 1 SD above them in skill but who lacked the specialized knowledge (for a similar approach using the interval Elo scale to quantify effects, see Bilalić, McLeod, & Gobet, 2008a, 2008b).

Additional evidence of the importance of context is provided by the result on the Neutral problems. The same players who showed superior recall performance on the problems from within and inferior performance with the problems outside their specialization, now, in a context equally familiar to both groups, displayed similar recall performance. The superior performance on the recall task on the positions within the area of specialization probably relies on more fine-tuned knowledge structures. Template theory (Gobet & Simon, 1996a) states that with time and extensive practice, knowledge becomes more and more complex and differentiated. The most complex, finely tuned knowledge structures, templates, vary between individuals as a function of exposure to certain types of stimuli. In chess, they can be a consequence of opening specialization. They are also a characteristic of higher levels of expertise and are almost exclusively found among highly skilled players (Gobet & Simon, 1996a). Slightly less-complex knowledge structures, chunks, on the other hand, are more common for weaker players, but there is evidence that even experts will have a number of similar chunks in common (Gobet & Simon, 1998). Given that all participants in this study were highly skilled chess players, it is reasonable to assume that the difference in the recall between the two groups of differently specialized experts was a direct outcome of the differences in the structure of their templates. When the players had to deal with the neutral middle-game positions for which both groups had a number of shared chunks, varying in complexity depending on the skill level, the differences between players of the same skill level but with different specializations disappeared.

Similar findings were found in the problem-solving part of the study. When confronted with problems within their area of specialization, they were more likely to generate more successful solutions to such an extent that their performance was similar to that of the players nominally stronger (for 1 SD) but solved the same problems outside their area of specialization. The skill differences, however, were intact on the Neutral problems belonging to neither of the two areas of specialization. These results imply that there is indeed a close link between knowledge/memory, as captured by the recall task, and problem-solving performance, as captured by the find-the-best-move task. Although this would not be surprising to chess experts who know the advantage of being in a familiar situation, the results have theoretical significance and practical implications.

The fact that the results on problem-solving mirrored those on memory recall supports Gobet and Simon's (1996a) hypothesis that templates are connected to potential solutions and plans. Just as the perceptual knowledge of chunks and templates grows with the exposure to opening specific stimuli within the domain, so does the knowledge of the possible actions that can be associated with them. Knowledge of perceptual patterns is of little use without knowing what methods should be used with them, but knowledge of methods is also insufficient for a high level of expertise without knowledge of the situations where these methods are likely to succeed. In our problem-solving task, the players performed best when they were dealing with problems from their domain of specialization, for which they had acquired both perceptual chunks/templates and knowledge of what actions, strategies, and tactics followed from their activation. Thus, as argued more generally by Gobet (2005) and Zhu, Lee, Simon, and Zhu (1996), becoming an expert requires both the accumulation of a large number of domain-specific patterns and the development of increasingly differentiated methods of action. These considerations have direct applied implications for education. For example, Zhu et al.'s research led to new and more efficient mathematics and physics curricula in China. Students were specifically encouraged to learn new perceptual chunks rather than to focus on the actions without knowing when they were appropriate.

How can we then reconcile our findings on the memory and problem-solving performance with the findings of training studies showing that superior memory is possible in the absence of superior problem-solving skills (Ericsson, 1985; Ericsson & Chase, 1982; Ericsson & Harris, 1990; Ericsson & Oliver, 1989). It is important to point out that even memory superiority of those individuals was based on their previous knowledge. The digit-span experts did not have extensive experience in the domain, but they used their previously acquired knowledge of dates and other numbers for new material. As we have noted above, chunk-based theories explain experts' performance by the assumption that they have to learn perceptual chunks, relevant actions, and links between chunks and actions. The chess-training experiments related only to the first of these three components of learning. The rapid improvement seen in two novices trained to memorize chess positions (Gobet & Jackson, 2002) are successfully modeled by CHREST (for a similar explanation, see also Ericsson & Kintsch, 1995; Ericsson & Lehmann, 1996).

The specialization effect was also evident in experts' problem-solving strategies. When confronted with problems within their area of specialization, they investigated fewer

solutions, spent less time, generated fewer episodes, and wandered less than players of the same level of skill who were outside their area of specialization but considered solutions in greater depth. The players who were more familiar with the positions concentrated on the most likely solutions and investigated them in more detail. It is not surprising that greater depths of search were associated with the problems within the opening of specialization and greater breadth with the positions outside. The problems outside the area of specialization were so unfamiliar that it sometimes was a real surprise for the players to be involved in solving such positions. For example, a Sicilian IM said, ''I think this is French Winawer. I do not know anything about this opening!'' when confronted with the first French position. Just a few positions later, the second French position occurred that elicited a reaction of an unpleasant surprise: ''Oh dear, another French Winawer!'' Probably the most telling reaction about what goes on when a problem solver is confronted with an unfamiliar problem is the following quote from the protocol of a French M who was tackling a Sicilian position: ''This is the Sicilian as well, I think. I do not play the Sicilian with White nor with Black so it is difficult for me to grasp the complete problematic in a short period of time. Well, I will have to try to do it using normal chess reasoning.'' He went on to analyze the position in great detail and eventually found the right path.

The Neutral problems again provide the evidence on the importance of the context. Suddenly, there was no difference in problem-solving strategies between the two types of players on Neutral problems. The same players who showed different patterns of search when confronted with problems from within and outside their specialization now, in a context equally familiar to both groups, used similar problem-solving strategies. The differences between skill levels in problem-solving strategies were also evident on the specialized problems in that super experts (IM&GM) were not only better at solving the problems, but they also used different search strategies from weaker experts (CM). The differences between skill levels in strategies, however, were in particular clear cut on the first Neutral problem. Problem-solving performance increased as the expertise increased, but the other protocol parameters also indicated that super experts used different strategies too. Most notably, their depth of search was noticeably greater than that of their weaker colleagues.

At first sight, this seems to contrast with the finding from de Groot (1978) and the main assumption in theories of expertise (e.g., Gobet, 1997, 1998a, 1998b) that super experts and ordinary experts use similar search strategies. It is possible that deep search was not necessary for the best solution to be discovered in the position that de Groot used to draw his conclusion.[5] Hence, super experts did not need to search deeply but they were able to select more pertinent candidate moves for search than experts. The depth of search depends on the need for it. If there is no need, as in de Groot's A position, super experts will not engage in deep search.

The problem-solving strategies seem to depend also on the difficulty of the problem. The differences that were observed on the specialization problems and the first neutral problem did not hold on the more difficult second neutral problem. Here, the problem-solving strategies, depth and breadth of search, were similar among differently skilled experts. Most surprisingly, super experts also did not find better solutions (although differences in both parameters were present favoring super experts). If familiarity with the problem and its

difficulty influence the search strategies experts employ, then the whole issue about experts' strategies seems less relevant for education. Problem-solving strategies will be a complex mix of individual characteristics (expertise) and the context (difficulty and familiarity of the problem). Strategies are a part of an expert's arsenal, but they are probably more of a product than a reason for the expertise. It is difficult to imagine that teaching someone the experts' strategies will result in a big increase in expertise. For example, searching in depth (also called forward reasoning), the "hallmark of expertise" (Chi et al., 1988; Patel & Groen, 1986, 1991; Patel et al., 1990) when not guided with extensive knowledge, inevitably leads to blind allies and wrong solutions as demonstrated in Eva, Brooks, and Norman's (2002) study. Finally, we saw that the strategies experts employ are flexible depending on the context which makes them difficult to pinpoint and teach. Different people, even if they are at the same high skill level, find different ways of dealing with problems that work well for them.

The conclusion of this study is the opposite of the conclusion of previous studies using the specialization paradigm (e.g., Schraagen, 1993; Schunn & Anderson, 1999; Voss et al., 1983) and especially of a similar assumption of the critical thinking movement (e.g., de Bono, 1982; Ennis, 1991, 1996). While there may indeed be some domain-general strategies which might be more efficient than very general methods, the so-called weak methods, their questionable value in the problem-solving process and education remains. Unspecialized experts (subexperts) did exhibit similar problem-solving process to the specialized experts, and their solving was highly structured, but that does not seem to make a difference. In both studies, which made qualitative predictions (Schraagen, 1993; Schunn & Anderson, 1999), unspecialized experts were outperformed by specialized experts while their solutions were not significantly better than that of undergraduate students who did not possess domain-general strategies. This lack of differentiation in the quality of solution between subexperts and novices casts doubt on the usefulness of domain-general strategies. It is more realistic to assume that knowledge (close to familiarity in this study) is more responsible for expertise than problem-solving strategies. It might be possible for weaker players to adopt the strategy of searching extensively and deeply. de Groot (1978) and this study show that ordinary experts, though they are skilled chess players, do not search the same solutions as super experts, although they might reach the same depths on the candidate moves they do search. The explorations focus of super experts was superior as evidenced by the quality of the solution. Search strategies must be directed by knowledge; otherwise it will be difficult to identify the relevant problem space for the correct solution that needs to be investigated.

Notes

1. Chess has an interval scale for measuring skill levels of chess players on the basis of their results against other players of known rating. The Elo scale has a theoretical mean of 1,500 and a theoretical *SD* of 200 (see Elo, 1978). Players with a rating between 2,000–2,200 are called Candidate Masters, 2,200–2,400 Masters, 2,400–2,500 International Masters, and above 2,500 Grand Masters. Candidate Masters (Elo 2,000–2,200)

are regularly called ''Experts'' in the expertise literature. However, in this paper, we used ''Candidate Masters'' to avoid confusion with the stronger players in our sample who are undeniably ''experts.''

2. One could question the use of Fritz to analyze the solutions. Fritz and chess computers, in general, are traditionally considered to be suitable for analyzing complicated ''tactical'' problems (for example, see Chabris & Hearst, 2003). At the same time, peaceful ''positional'' problems (such as those used in this experiment) are considered to be difficult for computers. The new-generation computers, however, are also able to successfully deal with positional problems. This was evident when a newer version of Fritz recently beat the world champion in a match. Furthermore, the evaluation from Fritz correlated highly ($r = .77$) with the evaluation of a Master (the first author) who used a 1–5 scale to evaluate the moves.

3. According to Fritz's evaluations, the Sicilian positions seemed to be easier than the French positions although the difference did not quite reach significance ($F[1, 2] = 6.99$, MSE $= 0.62$, $p = .12$, $\eta_p^2 = 0.78$). Thus, looking at the performance of individual groups on the two classes of problems, rather than the overall interaction, the effect may appear to be absent with the French players (that is, their Sicilian solutions are as good as their French). The correct comparison to use is between different players on the same problem rather between the same players on different problems.

4. The difficulty of this position, de Groot's position C, is illustrated by Campitelli and Gobet (2004). They let an IM and a GM investigate the position. After 30 min, much longer than would normally be taken by a player for a single move in a game, neither player reached the best solution.

5. Indeed, de Groot himself calculated that a search of only 5 ply was necessary to find the best solution.

Acknowledgments

References

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist*, *48*, 35–44.
Ball, L. J., Evans, J. S., Dennis, I., & Ormerod, T. C. (1997). Problem-solving strategies and expertise in engineering design. *Thinking and Reasoning*, *3*, 247–270.

Bilalić, M., McLeod, P., & Gobet, F. (2008a). Why good thoughts block better ones? The mechanism of the pernicious Einstellung effect. *Cognition*, *108*, 652–661.

Bilalić, M., McLeod, P., & Gobet, F. (2008b). Inflexibility of experts—reality or myth? *Cognitive Psychology*, *56*, 73–102.

Bilalić, M., McLeod, P., & Gobet, F. (2008c). Expert and 'Novice' problem solving strategies in chess—sixty years of citing de Groot (1946). *Thinking & Reasoning*, *14*, 395–408.

de Bono, E. (1982). *de Bono's thinking course*. New York: Facts on File.

Burns, B. D. (2004). The effects of speed on skilled chess performance. *Psychological Science*, *15*, 442–447.

Campitelli, G., & Gobet, F. (2004). Adaptive expert decision making: Skilled chessplayers search more and deeper. *Journal of the International Computer Games Association*, *27*, 209–216.

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.

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

Charness, N. (1989). Expertise in chess and bridge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 183–208). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Clarke, V. J., & Lamberts, K. (1997). Strategy shifts and expertise in solving transformation rule problems. *Thinking and Reasoning*, *3*, 271–290.

Elo, A. E. (1978). *The rating of chess players, past and present*. New York: Arco.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Ennis, R. H. (1991). Critical thinking: A streamlined conception. *Teaching Philosophy*, *14*, 5–24.

Ennis, R. H. (1996). *Critical thinking*. Upper Saddle River, NJ: Prentice-Hall.

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

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. *American Scientist*, *70*, 607–615.

Ericsson, K. A., & Harris, M. S. (1990, November). *Expert chess memory without chess knowledge: A training study*. Poster presented at the 31st Annual Meeting of the Psychonomic Society, New Orleans, LA.

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

Ericsson, K. A., & Kintsch, W. (2000). Shortcomings of generic retrieval structures with slots of the type that Gobet (1993) proposed and modelled. *British Journal of Psychology*, *91*, 571–590.

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

Ericsson, K. A., & Oliver, W. (1989). A methodology for assessing the detailed structure of memory skills. In A. Colley & J. Beech (Eds.), *The acquisition and performance of cognitive skills* (pp. 193–215). London: Wiley.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. Ed.). Cambridge, MA: The MIT Press.

Eva, K. W., Brooks, L. R., & Norman, G. R. (2002). Forward reasoning as a hallmark of expertise in medicine: Logical, psychological, phenomenological inconsistencies. In S. P. Shohov (Ed.), *Advances in psychology research* (Vol. 8, pp. 41–69). New York: Nova Science.

Gobet, F. (1997). A pattern-recognition theory of search in expert problem solving. *Thinking & Reasoning*, *3*, 291–313.

Gobet, F. (1998a). Expert memory: A comparison of four theories. *Cognition*, *66*, 115–152.

Gobet, F. (1998b). Chess players' thinking revisited. *Swiss Journal of Psychology*, *57*, 18–32.

Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology*, *19*, 183–204.

Gobet, F. (in press). Using a cognitive architecture for addressing the question of cognitive universals in cross-cultural psychology: The example of awalé. *Journal of Cross-Cultural Psychology*.

Gobet, F., & Jackson, S. (2002). In search of templates. *Cognitive Systems Research*, *3*, 35–44.

Gobet, F., & Simon, H. A. (1996a). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*, 1–40.

Gobet, F., & Simon, H. A. (1996b). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grandmaster level chess. *Psychological Science*, *7*, 52–55.

Gobet, F., & Simon, H. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, *6*, 225–255.

Gobet, F., & Waters, A. J. (2003). The role of constraints in expert memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *29*, 1082–1094.

de Groot, A. D. (1978). *Thought and choice in chess* (2nd Ed.) The Hague: Mouton Publishers (originally published 1946.)

de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess. Studies in the heuristics of the professional eye*. Assen: Van Gorcum.

Gruber, H. (1991). *Qualitative Aspekte von Expertise im Schach*. Aachen: Feenschach.

van Harreveld, F., Wagenmakers, E. J., & van der Maas, H. (2007). The effects of time pressure on chess skill: An investigation into fast and slow processes underlying expert performance. *Psychological Research*, *71*, 591–597.

Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Erlbaum.

Holding, D. H. (1992). Theories of chess skill. *Psychological Research*, *54*, 10–16.

Hunt, E. (1991). Some comments on the study of complexity. In R. J. Sternberg & P. A. Frey (Eds.), *Complex problem solving* (pp. 343–381). Hillsdale, NJ: Erlbaum.

Jeffries, R., Turner, A. A., Polson, P. G., & Atwood, M. (1981). The processes involved in designing software. In J. R. Anderson (ed.), *Cognitive skills and their acquisition* (pp. 255–283). Hillsdale, NJ: Erlbaum.

Joseph, G. M., & Patel, V. L. (1990). Domain knowledge and hypothesis generation in diagnostic reasoning. *Medical Decision Making*, *10*, 31–46.

Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine*, *13*, 110–116.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335–1342.

Linhares, A. (2005). An active symbols theory of chess intuition. *Minds and Machines*, *15*, 131–181.

Linhares, A., & Brum, P. (2007). Understanding our understanding of strategic scenarios: What role do chunks play. *Cognitive Science*, *31*, 989–1007.

Newell, A. (1980). Reasoning, problem-solving and decision processes: The problem space as a fundamental category. In R. Nickerson (Ed.), *Attention and performance, VIII* (pp. 693–718). Hillsdale, NJ: Erlbaum.

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

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, *10*, 91–116.

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: Prospects and limits* (pp. 93–125). Cambridge, England: Cambridge University Press.

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory and Cognition*, *18*, 394–40.

Reimann, P., & Chi, M. T. (1989). Human expertise. In K. J. Gilhooly (Ed.), *Human and machine problem solving* (pp. 161–191). New York: Plenum Press.

Saariluoma, P. (1990). Apperception and restructuring in chess players pro-blem solving. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), *Lines of thought: Reflections on the psychology of thinking* (Vol. II, pp. 41–57). New York: Wiley.

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

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

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

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, *23*, 337–370.

Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, *105*, 33–57.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (Vol. 1, pp. 205–232). New York: Academic.

Waters, A. J., & Gobet, F. (2008). Mental imagery and chunks: Empirical and computational findings. *Memory & Cognition*, *36*, 505–517.

Williams, M. W., Papierno, B. P., & Makel, C. M. (2004). Thinking like a scientist about real-world problems: The Cornell Institute for Research on Children Science Education Program. *Journal of Applied Developmental Psychology*, *25*, 107–126.

Zhu, X., Lee, Y., Simon, H. A., & Zhu, D. (1996). Cue recognition and cue elaboration in learning from examples. *Proceedings of the National Academy of Sciences*, *93*, 1346–1351.

Appendix

Factor analysis of protocol parameters

Factor analysis (FA) with Varimax rotation was conducted for French, Sicilian, and Neutral problems separately as well as for all positions together on the parameters extracted from the protocols of players. All FA produced two factors that can be broadly classified as factors of depth and breadth of search. The first factor in all analyses presented the parameters of search depth (mean depth and maximal depth), while the second factor always included the parameters of search breadth (candidate moves and episodes; see Table A).

Table A

Factors and their loadings for protocol parameters for French, Sicilian, Neutral, and all positions together

| | Factors | | | | | | | |
| | French Positions | | Sicilian Positions | | Neutral Positions | | All | |
Parameters	Depth	Breadth	Depth	Breadth	Depth	Breadth	Depth	Breadth
Total nodes	0.52	**0.82**	**0.97**		0.43	**0.88**	0.59	**0.78**
Mean depth	**0.91**		**0.81**	0.39	**0.93**		**0.90**	
Maximal depth	**0.89**		**0.85**		**0.86**		**0.87**	
Candidates		**0.91**		**0.80**	−0.50	**0.67**		**0.82**
Episodes		**0.99**	0.39	**0.90**		**0.96**		**0.97**
Branches	0.61		**0.76**		0.52	0.74	0.66	0.49
Reinvestigations		**0.88**	**0.93**			**0.87**		**0.82**

Note: Factor loadings <0.30 were not presented. Bold indicates belonging to one of the factors.

Number of branches and immediate reinvestigations were frequently found on the depth factor, while reinvestigations and total moves were mostly connected with the breadth factor. These measures did not show a consistent connection to either of the two factors across different position types. Therefore, we used only mean and maximal depth of search as the indicators of depth of search and number of candidates and episodes as the indicators of breadth of search.