ROLES OF PATTERN RECOGNITION AND SEARCH IN EXPERT PROBLEM SOLVING

FERNAND GOBET
ESRC Centre for Research in Development, Instruction and Training
Department of Psychology, University of Nottingham
University Park, Nottingham NG7 2RD
United Kingdom
email: frg@psyc.nott.ac.uk

ESRC Centre for Research in Development, Instruction and Training
August 1996
Technical Report No. 42

Abstract
Understanding how search and pattern recognition interact is one of the important research questions in the study of expert problem-solving behavior. This paper examines the implications of a recent theory of expert memory, the template theory (Gobet & Simon, in press-a), on the theory of problem solving. Templates are chunks (Chase & Simon, 1973) that have evolved into more complex data structures, allowing values to be encoded into slots rapidly. It is proposed that templates facilitate search in three ways: (a) by allowing information to be stored into LTM rapidly; (b) by allowing a search in the template space in addition to a search in the move space; and (c) by compensating loss in the mind's eye due to interference and decay. The template theory accounts for the slight skill difference in average depth of search found in chess players, as well as for other empirical data.

Email: frg@psychology.nottingham.ac.uk
Phone (0115) 951 5402, Fax (0115) 951 5324

1. INTRODUCTION

How much expert problem-solving behavior is explained by real-time search through the task problem space, and how much is explained by pattern recognition? Ever since the seminal work of De Groot (1946/1978), this question has been a recurring theme in the study of expertise (see Charness, 1991; Holding, 1985).

The study of chess players, originated in its modern form by De Groot (1946), has been a productive subfield of the psychology of expertise. Chess is a task where perception, memory, knowledge organization and search interact in a complex, dynamic way. The advantages it offers include: a quantitative scale for measuring skill, a large database of games, and ecological validity of the domain. In addition, numerous empirical data have been collected and several psychological theories of chess skill have been proposed. Taken together, these features explain the attention cognitive scientists have paid to chess.

The goal of this paper is to extend the template theory (Gobet & Simon, in press-b), which was originally intended as a theory of chess players' perception and memory, to the realm of problem solving. After having proposed elements of a comprehensive theory of chess problem solving that is compatible with empirical data on memory and perception research, I will briefly see how it generalizes to other domains of expertise.

2. CHESS PROBLEM SOLVING: THEORIES AND DATA

The basic finding of De Groot's study (1946/1978) was that there were no important differences in the macrostructure of search between world-class grandmasters and strong amateurs, except that the best players were somewhat faster to reach a decision and that they found better moves. On average, players from both skill levels searched equally deep, visited the same number of positions in their search, and proposed the same number of candidate moves. De Groot concluded that the key to expertise is not a superiority in "speed of thinking" or in a more rational way to organize one's thought, but lies in specific knowledge about various aspects of the task domain. De Groot noted that the search behavior shown by his subjects was compatible with the general theoretical background proposed by Selz (1922) but he did not offer a detailed model to account for their behavior.

Such a detailed account was offered by Chase and Simon (1973) and Newell and Simon (1972). Their theory, commonly known as the chunking theory (CT), proposed that at the core of expertise lies the ability to rapidly recognize important features in the problem at hand. These features, internally stored as chunks, act as access points to semantic long-term memory (LTM) and as the conditions of productions, whose actions may be carry out internally or externally. This production-system account was linked to assumptions about information-processing mechanisms, describing for example how an object is sorted through a discrimination net in order to reach a node in LTM. It also included parameters specifying learning and accessing times (e.g., about 8 s to learn a new chunk), as well as capacity limits (e.g., STM limited to 7 items). In the case of chess expertise, perception mechanisms in CT allow recognition of patterns of pieces on the board. These patterns suggest moves, which are used to update the internal representation of the board in the mind's eye. The mind's eye is a relational system that can be subjected to visuo-spatial mental operations and that stores perceptual structures, both from external inputs and from memory stores. Recognition mechanisms apply recursively in the positions as represented in the mind's eye. Termination of search in a branch is obtained by evaluating whether certain goals are above or below a threshold, whose value may change as a function of players' aspiration level of the position.

Holding (1985, 1992) systematically criticized CT, and proposed instead the SEEK (SEarch, Evaluation and Knowledge) theory, a theory that was however only loosely framed and did not reach the theoretical detail of CT. Note that all three components of the SEEK theory are present in CT and that it is really the weight given to each of these components that differentiate the two theories. In particular, Holding proposed that strong players search much deeper and wider in the tree search than was suggested by De Groot (1946).

Recent empirical data support Holding's contention that stronger chess players search deeper, on average, than weaker players (Charness, 1981; Gobet, 1995), but not that they search wider. It is however unclear to what extent the differences in average depth of search support Holding's claim that they refute the chunking hypothesis: these differences are small, and, as discussed below, CT also predicts that strong players have a higher ability to search, even if they may not need to do so in most cases.

Recently, Saariluoma (1990) proposed that players, while trying to find a move in a position, accessed goal-positions by pattern recognition (Saariluoma uses the term "apperception"), and then tried to close the path from the problem position to the goal-position. I will propose later than this path-closing behavior may be seen as a special case of the more general theory outlined in this paper.

Finally, Ericsson and Kintsch (1995) recently proposed, within their long-term working memory theory (LTWMT), that skilled players have constructed a hierarchical retrieval structure, not unlike the structure created by SF and DD, two mnemonists specialized in the digit-span task (Chase & Ericsson, 1982). This powerful memory structure, isomorphic to the 64 squares of the chess board, allows players to update information rapidly during search. I have shown elsewhere (Gobet, 1996-a) that LTWMT generates predictions that do not agree with the experimental data on perception and memory. In addition, the retrieval structure proposed by LTWMT seems at odds with the following aspect of chess players' search behavior. It is common for chess players to revisit nodes during their analysis of the position. Now, instead of accessing this node directly, as suggested by LTWMT's very powerful mechanism for storing positions rapidly and accurately, players access the target position by generating individual moves from the problem position. This behavior suggests that players are not able to store positions as fast as proposed by LTWMT.[1]

3. THE TEMPLATE THEORY AND ITS APPLICATION TO PROBLEM SOLVING

Initially, the template theory (TT) was developed to remove some theoretical weaknesses in the chunking theory with respect to data from memory research (Gobet & Simon, in press-b). In particular, contrary to CT predictions, chess players' memory is resistant to interference (Charness, 1976) and chess players can recall several briefly-presented positions with reasonable accuracy (Cooke, Atlas, Lane & Berger, 1993; Gobet & Simon, in press-b).

Like CT, TT proposes that chunks are accessed by traversing a discrimination net. Chunks are linked to other information stored in LTM, such as moves, plans, tactical motives, and so on. In addition, chunks that often occur in a player's practice and study evolve into more complex data structures (templates), which have slots allowing variables to be encoded rapidly, for example, to specify the square occupied by a piece. In addition to slots, templates contain a core, which is basically similar to the information stored in chunks. Altogether, templates are assumed to store around ten pieces or more. Finally, templates may be linked to other templates. TT also proposes that the information stored in the mind's eye decays rapidly, and that it needs to be updated either by inputs from the external world or by inputs from memory structures.

TT removes the difficulties of CT with interference results, first because there is some fast LTM encoding (through template slots) and second because templates contain more information than the chunks proposed by CT (CT assumed that chunks contain at most four or five pieces). Finally, TT keeps the power of CT to explain the rapid recognition of positions shown by masters, since the same discrimination mechanisms are present (Gobet, 1996-a, 1996-b).

3.1 Search mechanisms

TT, as does CT, proposes that search is carried by recursively applying pattern recognition processes in the internal representation. In the remainder of this section, I will flesh out this mechanism with additional specifications. I will also show how chunks and templates may be used to terminate search, and how templates may be used to carry out search at a more abstract level than at the move level.

3.1.1 Move generation

As mentioned above, pattern recognition is applied recursively. If a chunk[2] is recognized, two cases may occur. First, the chunk gives access to a move or a sequence of moves. In this case, the move(s) is (are) carried out in the mind's eye. Second, the chunk gives access to heuristics for selecting the next move, such as: "Counter-attack in the center when attacked on the King's side." In this case, the heuristics are applied, and a move or a sequence of moves is possibly carried out. If no move is yielded by either of these mechanisms, one of two operations may be applied. Either there is an attempt to access, through semantic links, a chunk that yields information about move(s) or possible heuristics, or weaker heuristics are applied, such as "occupy the center" or "attack the opposite king."

3.1.2 Update of internal representation and storage of information into template

When a move or a sequence of moves is carried out in the mind's eye, the internal representation is updated, and, when a template has been accessed, slots are updated as well. Slots may store information about the location of pieces, strategic information, evaluation of the position or part of it, and so on. Note that this information facilitates search, because possible forgetting through decay in the mind's eye may be backed-up by information stored in slots.

3.1.3 Evaluation of nodes

Evaluation of the position is not carried out after each move—this would be too expensive from an computational point of view—but only at the terminal positions. Evaluation either is retrieved, when possible, from chunks that have been recognized, or is computed when the player decides to terminate search (see next paragraph). De Groot's (1946) protocols indicate that players' evaluations are rather simple, taking into account a single feature, and strongly suggest that evaluations are not computed by conjoining several features, such as mobility or control of the center.

3.1.4 Ending search of a branch

A critical question, given a limited information-processing capacity, is when to stop searching a branch. I propose that the following conditions can induce halting search of a branch. To begin with, the risk of making an error is deemed higher than some parameter, which may vary across players. Psychologically, this probability of making an error is a function of the forgetting and interference rates in the mind's eye. Phenomenologically, players probably assess this risk by taking into account the depth of search reached and by estimating their familiarity with the position.

Another reason to halt search occurs when the aspiration level of the position (Newell & Simon, 1972) is met or when a feature of the position (e.g., safety of the king) was evaluated to be lower than its admissible value. This aspiration level was provided by a chunk met earlier during search or by explicit evaluation of the position, and was stored either in STM or, when available, in a template slot. Finally, some decisive circumstance occurred such as mate, important material gain or loss, or fulfillment of a key goal.

3.1.5 Selecting the next branch

We consider now the question of which base-move to search next when a leaf node has been reached. Six mechanisms may occur. First, if the leaf was evaluated with a high risk of error, the same line is searched again in order to obtain a more reliable estimate. Second, the same line may be searched again to make sure that it has been stored in LTM. Third, if some template has been recognized by pattern recognition or by following semantic links, a meta-search (see below) is carried out. Fourth, there is a bias to give preference to normal base-moves early in search, and to unusual base-moves later (Newell & Simon, 1972, p. 723). Fifth, a move reached at any level of depth in searching a line may be tried as a candidate move in the next line. Sixth, if the evaluation of an episode gives a favorable result, its analysis is continued; if the evaluation is unfavorable a different base-move is taken up (Newell & Simon, 1972, p. 723).

3.1.6 Meta-search with templates

Search may be carried out not only at the move level, but also at the template level. By this, it is meant that players sometimes develop a tree search by generating key positions (templates) instead of developing a tree search by generating moves. Paths between key-positions consist in unspecified operators, which will be filled in at a later stage of search, if the leaf positions are judged favorable. This type of search is contingent to the presence, in players' LTM, of templates sufficiently close to the position at hand. This meta-search mechanism is similar to the "closing-path" mechanism described by Saariluoma (1992), which was, however, left largely unspecified.

3.2 Theoretical importance of templates

Obviously, many of the processes described above can be implemented within the earlier CT, without hypothesizing the presence of templates. Templates are, however, important for some of the key processes, because (a) slots allow information about strategic or tactical themes, as well as about evaluation of the position, to be encoded in LTM rapidly; (b) links between templates allow for "meta-search;" (c) information stored both in the core and in the slots of the template compensates decay and interference in the mind's eye, as it may be used to reinstate fading information (this is particularly useful when perceptual information from the external board is unreliable, for example because several moves have been carried out in the mind's eye).

3.3 Empirical data supporting the TT account of search and pattern recognition

While it is beyond the scope of this paper to systematically confront TT with empirical evidence from problem solving research on chess, it may be useful to see how the theory fares with some of the key data of the domain. (I have shown elsewhere [Gobet, 1996-b] that the predictions of the theory are well supported by empirical data from research on memory and perception.)

• Growing the number of chunks and templates has two opposite effects with respect to search (Gobet, 1995). On the one hand, deep search is facilitated, as compared with the more computationally expensive weak methods, because pattern recognition is a rapid process and because sequences of moves may be suggested by chunks. On the other hand, the need of deep search often disappears because evaluations suggested by pattern recognition may be more reliable than evaluations computed by weak

methods. These two opposite effects may explain why the difference in average depth of search is small.
• Pattern recognition allows strong players to be highly selective in their search. This can be seen both in the rapid way masters zero in into the key aspects of a position (De Groot, 1946), and in the way eye movements differ as a function of skill during the first seconds of the perception of a position (De Groot & Gobet, in press). Additional evidence is provided by Saariluoma (1990), who has shown that players choose stereotyped solutions in problems where faster (in the number of moves to reach mate) but uncommon solutions are present. These data add support to TT, because pattern recognition lies at its core.
• Masters playing several opponents simultaneously tend to play "normal" moves and rely on opponents' errors (Gobet & Simon, in press-a). Normal moves may be defined as the type of moves suggested by templates.
• De Groot (1946) has proposed that players organize positions around large "complexes," bigger than the chunks that were later proposed by Chase and Simon (1973). De Groot also suggested, as did more recently Freyhoff, Gruber and Ziegler (1992), as well as Cooke et al. (1994), that players represent information at a conceptual level. TT offers theoretical structures (templates) that account for players' large complexes and high-level representations. TT also offers mechanisms explaining how these high-level representations are developed and how they are accessed rapidly.
• Holding and Reynolds (1982) briefly presented subjects with semi-random positions and found no skill difference in recall. Next, subjects had to find a good move in these positions. In this task, there was a strong correlation between skill level and quality of move proposed. Holding and Reynolds conclude that, because players can find better moves even when they cannot recognize more chunks, theories based on pattern recognition must be incorrect. Their conclusion is unwarranted, however, as can be seen from the following explanation. Assume that players recursively generate a move through weak methods when no chunk is recognized, until one of the halting conditions mentioned earlier applies. Then, given their larger database of chunks, strong players are more likely than weak players to find a useful pattern in positions that have been newly updated in the mind's eye, and therefore are more likely to find good moves in their search.
• We saw earlier that Ericsson and Kintsch (1995; see above) proposed a retrieval structure that encodes information fast and reliably. By contrast, the encoding capacity of templates is relatively weak. In particular, TT predicts that the same template may be used to encode several similar positions, thus increasing the risk of interference as search progresses. The consequence is that, contrary to LTWMT's predictions, players should reach a previously visited position by generating moves from the initial position, instead of accessing it directly by relying on their memory. This what players—even masters—do (De Groot, 1946).
• Krogius (1976) mentions that errors often occur when a move has caused important changes in the position. This may be due to players' inability to recognize that a template previously adequate has become inadequate.
• Masters' descriptions of games are centered around key positions of the game (cf. De Groot's [1946] protocols). I suggest that some of these key positions refer to templates.

4. CONCLUSION

Previous research has identified some crucial components of expertise: ability to perceive the material at several levels (chunks and high-level representations) and excellent memory in the task domain, on the one hand, and ability to search, on the other hand. Instead of opposing these components in a yes-or-no fashion, as it has been done sometime, for example by Holding (1985), it seems more important to understand the relations between these components. The theory developed in this paper has offered mechanisms for two important aspects of chess problem solving:

• Search is highly directed by memory and perceptual processes. At the same time, search will determine what will be perceived in the mind's eye.

• There is an interplay of high-level representations of the problem with low-level representations.

To what extent do the theoretical ideas put forward in this paper generalize to other domains of expertise? One of the main teachings of recent research is that different task environments tax cognitive functions differently. In particular, environments differ massively in the amount of search needed to carry out the task successfully. Hence, it is dubious that findings from chess research will apply without qualification to other domains. However, the interactions between low-level and high-level representation, on the one hand, and between knowledge and search, on the other hand, are also apparent in several domains, such as medical diagnosis, physics and text comprehension. The template theory, which accounts for these interactions, thus offers the perspective to integrate these various domains of expertise into a single framework.

5. REFERENCE LIST

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 641-653.

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 467-476.

Charness, N. (1991). Expertise in chess: the balance between knowledge and search. In K. A. Ericsson & J. Smith (Eds.), *Studies of expertise: Prospects and limits*. Cambridge: Cambridge University Press.

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

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Cooke, N. J., Atlas, R. S., Lane, D. M., & Berger, R. C. (1993). Role of high-level knowledge in memory for chess positions. *American Journal of Psychology, 106*, 321-351.

De Groot, A. D. (1946). *Het denken van den schaker*. Amsterdam: Noord Hollandsche.

De Groot, A. D. (1978). *Thought and choice in chess*. (Revised translation of De Groot, 1946; 2nd ed.). The Hague: Mouton Publishers.

De Groot, A. D., & Gobet, F. (in press). *Perception and memory in chess: Heuristics of the professional eye.* Assen (NL): Van Gorcum.

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

Freyhoff, H., Gruber, H., & Ziegler, A. (1992). Expertise and hierarchical knowledge representation in chess. *Psychological Research, 54*, 32-37.

Gobet, F. (1995). *Chess thinking revisited.* Submitted for publication.

Gobet, F. (1996-a). *Long-term working memory: Comment on Ericsson and Kintsch (1995).* Submitted for publication.

Gobet, F. (1996-b). *Memory in chess players: Comparison of four theories.* Submitted for publication.

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

Gobet, F., & Simon, H. (in press-b). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*.

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.
Holding, D. H., & Reynolds, R. I. (1982). Recall or evaluation of chess positions as determinants of chess skill. *Memory & Cognition, 10*, 237-242.
Krogius, N. (1976). *Psychology in chess*. London: R.H.M Press.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
Saariluoma, P. (1990). Apperception and restructuring in chess players' problem solving. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), *Lines of thinking* (Vol. 2). New York: Wiley.
Selz, O. (1922). *Zur Psychologie des produktiven Denkens und des Irrtums.* Bonn: Friedrich Cohen.

ACKNOWLEDGEMENTS

I thank Julian Pine and Frank Ritter for comments on a first draft of this paper.