Chunks and Templates in Semantic Long-term Memory:
The Importance of Specialization

Fernand Gobet

(11,317 words)

Abstract

The papers that Bill Chase wrote on skill in chess (with Herb Simon) and in the digit-span task (with Anders Ericsson) count among the most influential publications in the literature on expertise. Yet, the data discussed in these papers are based on the results of very few subjects, whose performance is essentially analyzed as case studies. In this chapter, I argue that case studies and individual data analysis can provide powerful information for developing and testing theories. I describe a 2-year study that, in the spirit of Chase and Ericsson (1981), followed the improvement of a chess master trained to memorize as many chess positions as possible, each presented for 8 seconds. The data are analyzed with a focus on the question of specialization of expert knowledge, and are compared to computer simulations carried out with the CHREST cognitive architecture. The question of specialization is further investigated with recent data that focus on chess players specializing in different openings. In line with Chase and Simon's chunking theory, CHREST, and other theories based on the concept of chunking, the results show a clear effect of specialization. Implications are drawn not only for the study of expert behaviour but also for the very applied question as to whether the quality of scientific papers can be evaluated reliably.

Gobet, F. (in press). Chunks and templates in semantic
long-term memory: The importance of specialization.
In J. Staszewski (Ed.), Expertise and skill acquisition:
The impact of William G. Chase. New York, NY: Psychology Press.

Introduction

This chapter will weave together several themes, focusing on two main questions. The first question relates to the use of single-subject designs in psychology. The second question pertains to the role of specialization of knowledge in experts. Specifically, it assesses the relative contributions of perceptual knowledge (e.g., chunks) and general knowledge (heuristics and strategies) to expertise. These two questions are interrelated because probing experts' knowledge is a complex and intensive task, and single-subject designs are one possible avenue for addressing this issue.

My views on the two questions are deeply influenced by Bill Chase: he carried out highly influential studies investigating experts' knowledge with the use of single-subject designs, and the main theoretical tools used in this chapter directly derive from the research he carried out on chess with Herb Simon.

The article starts with a discussion of multiple-subject designs, which clearly dominate research in psychology. After having discussed some limits of such designs, I will consider single-subject designs and highlight their advantages. Strengths and weaknesses of experimental designs resonate with many important questions in the field of cognitive modeling, which can be summarized as "What are the best experimental designs for developing cognitive models?" I will build on a proposal offered by Gobet and Ritter (2000) arguing that single-subject design can provide a powerful tool for developing and testing computational models in psychology. This will be illustrated by an experiment on chess memory, in which a single participant was trained to memorize as many briefly presented chess positions as possible. The analysis will focus on the effect of specialization in this master's performance, and computer simulations will show that the CHREST architecture captures important aspects of the empirical results. I will then show that these results generalize well, by discussing an experiment in which the question of specialization in chess was studied with a larger sample, with respect to both memory and problem solving. The final section of this chapter will draw implications of this research for the question of specialization in science, taking as example the recent Research Assessment Exercise (RAE) in England. The general conclusion is that experts' ability to successfully transfer their skills beyond their domain of specialization is more limited than generally thought.

Multiple-Subject and Single-Subject Designs

Research in cognitive psychology (and psychology in general) is predominantly done with groups of participants. This is also the case in expertise research, where it is difficult to publish results if the study does not contain several participants for each skill level. There is no doubt that this methodology is powerful, in particular due to the increased availability of sophisticated statistical techniques to make sense of the data. However, this methodology does have its undeniable limitations, as was noted by Newell and Simon (1972) and Siegler (1987), among others. For example, data averaged across people may not accurately reflect the behavior of any one person; participants may use different strategies and have different knowledge bases; and, with respect to the study of expertise, experts specialize in different sub-fields.

Single-subject designs are an obvious way to get around these limitations. Such a methodology has long history in psychology. Ebbinghaus (1988/1964) single-handedly created the field of memory research by using himself as his only subject. In their seminal study on telegraphers learning Morse, Bryan and Harter (1899) studied only two participants – but in great detail. Although De Groot's (1965) classic work on chess included a number of players as participants, his analyses were in large part qualitative and non-statistical – essentially a sequence of single-subject studies. And, of course, some of Bill Chase's most influential studies were carried out with very few participants. His study on chess skill (Chase & Simon, 1973a, 1973b; Simon & Chase, 1973) had only three participants (and only one per skill level). Similarly, his research on the digit span task focused on the study of a single participant, SF (Chase & Ericsson, 1981, 1982; Ericsson, Chase, & Faloon, 1980).

Indeed, Chase and Ericsson's work on the digit-span task offers a beautiful example of the power of single-subject designs. During the 25 months that the study lasted, they carried out a number of experiments aimed at testing the theory they had developed based on verbal protocols. One experiment tested the hypothesis that digit sequences difficult to encode as running times should lead to a poor performance. Another tested the converse hypothesis that digit sequences that fit SF's coding strategy should lead to an increase in performance. A third experiment tested the reliability of SF's encoding rules; this was done by presenting again a sequence that had been used one month earlier. Further experiments investigated the role of short-term memory in SF's behavior, by using several rehearsal-suppression techniques. Other experiments measured SF's short-term memory capacity and the possibility of transfer of his digit memory skills to other material. Together with the practice experiment itself and the associated verbal protocols, these experiments produced a uniquely rich set of data.

As illustrated by the research on digit span, single-subject designs offer a number of advantages and nice features (Campitelli, Gobet, Williams, & Parker, 2007; Gobet & Ritter, 2000). There is a rapid interplay between data collection and theory building, and, depending on the research design, new hypotheses can be tested quickly. Specifically, the results of one session lead to a new hypothesis, which can be tested in the next session. These features make single-subject designs particularly suitable for expertise research, as experimental effects tend to be large in this field of research and can be detected with a very small sample size – including one subject. However, in spite of these features, single-subject designs have a bad reputation in psychology. They have been criticized for lack of generalizability, lack of a control group, carry-over and order effects across tasks, and development of idiosyncratic strategies by the participant, who becomes an expert in participating in an experiment.

What Kind of Data to Use for Developing Cognitive Models?

Most models in cognitive psychology are tested with averaged data. This of course raises the same type of issues as those just discussed: average data may not reflect the behavior of any specific participant, and the fact that participants can have different strategies and knowledge bases is ignored. Most models in cognitive psychology also have multiple free parameters, an obvious weakness as it is often not clear whether the success of a model in simulating the data is due to its mechanisms or rather to its free parameters. One way around this problem, proposed by Newell (1990), is to develop Unified Theories of Cognition (UTCs). Here, a single architecture is to be used to account for as many empirical data as possible. According to Newell (and I agree with him), this allows one to limit the number of free parameters. However, this only addresses part of the problem: if the data to simulate are data averaged over, say, fifty participants, we still do not know whether they represent the actual behavior of any participant, and we still do not know the role played by differential strategy use and different levels of experience and knowledge.

Combining Single-Subject Designs and Cognitive Modeling

Gobet and Ritter (2000) have argued that combining single-subject designs with cognitive modeling makes the best of the two approaches. The idea is to gather a large number of empirical/experimental observations on a single participant (or a few participants analyzed individually), and then to develop a detailed computational model of each participant. The data should be detailed and varied enough – Chase and Ericsson's research on the digit span offers a good example of this – and, ideally, various task domains should be used.

Gobet and Ritter's (2000) methodology is not antagonistic to using group summaries. However, rather than using the observed data (e.g., percentage correct, reaction times, etc.) for computing aggregate statistical parameters such as the mean, they suggest first to estimate UTC-parameters for each subject individually, and then to calculate aggregate values over these parameters. This methodology for estimating aggregate values, which they call "between-subject analysis of parameters," is summarized in Table 1. According to Gobet and Ritter, this methodology can provide more robust and theoretically more meaningful estimates than the standard method for aggregating data.

It is important to emphasize that UTC-parameters are not limited to numeric values. For example, they can correspond to strategies. Possible strategies can be represented with a probability distribution, and, if one wants a single value, one could select the most common strategy (modal strategy). Siegler's (1987) work on children's addition strategies offers a good example of this approach. Incidentally, this highlights the dangers of averaging across participants, as done traditionally. For example, participants in a given task may use strategies that do not overlap, which makes averaging strategies meaningless. Across different trials with a same task and across several different tasks, participants are likely to show a pattern that clearly identifies their most common strategies.

-- Insert Table 1 around here --

Among other things, a benefit of this approach is that systematic sources of between-subject variability are illuminated rather than obscured: strategies and knowledge can be systematically studied rather than being ignored as random variation; learning is not noise but something to model and explain; and, if one is interested in expertise, the effects of specialization of knowledge can be studied. Thus, the "between-subject analysis of parameters" offers a powerful complement to group data analysis.

Specialization in Chess: A Single-Subject Training Experiment on Memorizing Multiple Chess Positions

Theoretical Motivation

The main aim of this experiment was to collect data to help revise Chase and Simon's (1973a) chunking theory. This theory was developed to explain the two key findings of De Groot's (1965) seminal research. First, chess masters can identify potentially good moves rapidly, often after only a few seconds. Second, they have an excellent memory for chess positions, even after a presentation as brief as 5 seconds. Chase and Simon replicated and extended De Groot's memory experiment. Based on an analysis of the latencies between the placements of pieces in a copy and a recall task, they proposed that masters encode information using *chunks.* Chunking had of course been well known in cognitive psychology since Miller's (1956) paper, but while Miller emphasized chunking as a strategic device to recode information in short-term memory (STM), Chase and Simon were interested in how chunks encode information in long-term memory (LTM). In this respect, chunks shared some resemblance to the "knowledge complexes" that De Groot had proposed, although he regarded the latter are more dynamic than the former. Chase and Simon argued that, compared to weaker players, chess masters have both more chunks and larger chunks in LTM. These chunks, whose maximum size was estimated at five or six pieces, encode information such as typical castle formations, pawn chains, common configurations on the first or eighth rank, and typical attacking patterns. Their data suggested that more than half of these chunks were pawn structures, which tend to remain relatively stable during a game.

To explain the roles of imagery and problem-solving in chess, Chase and Simon (1973b) proposed that LTM chunks are associated with processes that make it possible for the patterns to be reconstructed and manipulated as an image in the mind's eye. In addition, the chunks automatically activated by the patterns on the external chessboard trigger potential moves, which will be then further investigated by look-ahead search. What we have here is in effect a production system (Newell & Simon, 1972). It is important to stress that, in this theory, selecting a move involves not only pattern recognition, but also selective search.

This theory did a good job at explaining how the perceptual patterns that players acquire not only help them to memorize positions rapidly but also to find good moves in novel positions. It also provided a plausible explanation as to why it takes a long time – Chase and Simon proposed 10 years – to become a top-level player. With about 8 seconds to create a chunk (Simon, 1969), time is needed to acquire the 50,000 chunks that are necessary for attaining expert level. In addition to this, time is also necessary to learn the actions to carry out given a specific chunk, and to learn to pairs chunks with these actions. As noted by Richman et al. (1996), provision must also be made for relearning and over-learning information in order to compensate for the negative effects of forgetting, and for the fact that opportunity for learning novel information decreases as expert levels increase.

However, the chunking theory had two main weaknesses. First, as was shown shortly after its publication, the theory overestimates the time of encoding information in long-term memory (Charness, 1976; Frey & Adesman, 1976). Second, the theory underestimates the role of high-level knowledge: chunks are supposed to be fairly small (5-6 pieces at most), but evidence from verbal protocols clearly indicates that players use structures that are much larger. In fact, in some cases, the structures they use refer to the entire position (Cooke, Atlas, Lane, & Berger, 1993; De Groot, 1965; De Groot & Gobet, 1996). Although it is not strictly speaking a theoretical weakness, another issue should be mentioned here. Simon and Gilmartin's (1973) computer program, which simulated aspects of the Chase and Simon chunking theory, did not learn chunks autonomously but used chunks pre-selected by the modelers, and was not able to reach the recall performance of master level.

To correct these weaknesses, while still keeping the strengths of chunking theory, together with Herb Simon I proposed what we called the template theory (Gobet & Simon, 1996d). Chunks are still important in the theory, but are also complemented by more powerful memory structures – templates. Templates are schemas, with a core that encodes stable information and slots that encode variable information. They are created with chunks that are recognized often in a given domain. The originality of our proposal was that it assumed that information can be encoded rapidly (250 ms) in the template slots. Templates are also assumed to be linked to useful information, such as possible moves, evaluations, likely plans, and so on.

Another important aspect of templates is that they can be linked to other templates, either as a function of their similarity or a function of the temporal order they would normally occur in a game. For example, a template coding for a class of positions in the Scheveningen variation of the Sicilian defense (a common chess opening)[1] might be linked to typical middle game positions, which in turn can be linked to likely endgame positions. This makes it possible for chess experts to carry out search at an abstract level, and not only at the move level. Thus, the search would incorporate key positions and the focus would be on broad plans that could bring the player from one favorable key position to another. The detail of the exact sequence of moves is left for later analysis. Observation of discussions between chess players and perusing of commented games in chess books and magazines offer clear, albeit anecdotal, evidence that chess players carry out this kind of macro-level search. Templates can be seen as a type of retrieval structure, although they differ from the kinds of retrieval structures proposed by Chase and Ericsson (1982) and Ericsson and Kintsch (1995); see Ericsson and Kintsch (2000), Gobet (2000), and Gobet and Simon (1996d) for a discussion. I will have more to say about templates below, when presenting the computer model used for the simulations.

Thus, template theory combines low-level, perceptual memory structures with high-level and more abstract memory structures. The presence of templates explains

[1] In chess, the term “opening” refers to the first moves of the game. Over the years, an extensive body of knowledge – called “opening theory” – has developed about openings, and tens of thousands of books have been devoted to them. The length of theoretical variations varies from just a few ply to forty or even fifty ply for popular variations that have been extensively analyzed.

why chess masters use larger chunks that those proposed by the chunking theory.[2] It also explains why information is stored in LTM more rapidly than proposed by Chase and Simon. An important piece of information supporting both assertions was provided by experiments where players had to memorize not only one position at a time, as in the classic experiments by De Groot (1965) and Chase and Simon (1973a), but several positions presented in rapid sequence. The results obtained by Cooke, Atlas, Lane and Berger (1993) and Gobet and Simon (1996d) clearly established that chess players could encode much more information than what could be stored in seven STM chunks, assuming a maximal size of 5-6 pieces, as Chase and Simon did. This result is reminiscent of the semantic chunks for digit groups that SF and DD were able to encode and recall in the digit-span task (Chase & Ericsson, 1982; Staszewski, 1990).

The aim of the experiment, then, was to probe the limits of expert memory further, using the technique of presenting several chessboards. The methodology of this experiment was inspired from Ericsson and Chase's (1981), but uses chess as task domain rather than the digit-span task. The task consisted of remembering as many positions as possible, with each position being presented for 8 seconds. (A preliminary report of this experiment was provided in Gobet & Simon, 1996d.)

Method

[2] The reader might wonder why evidence for templates did not show up in the Chase and Simon's experiment. As established by further research using a computer program to present the positions and record the players' placements (Gobet & Simon, 1998), the limited capacity of the hand to hold chess pieces has led to an underestimate of the units encoded. Of particular interest in this respect is Gobet and Clarkson's (2004) experiment, in which the same participants carried out the experiment both with computer display and physical pieces. The size of the chunks was much larger in the former case.

A single participant P (the author of this chapter) took part in this experiment, which lasted about two years (213 sessions), with some interruptions due to holiday or work commitments. A former chess professional who then trained as a researcher in psychology, P was an International Master. At the time of the experiment, he had an USCF rating of 2396 Elo, which put him among the best 250 players in the United States.[3] However, he had barely played any competitive game for the four years before the experiment, and totally stopped practicing. During the experiment, P played little chess.

P was exposed to a number of positions and then afterwards attempted to reconstruct them on empty chessboards. The positions were randomly selected from a large database of games, and were taken after Black's 20th move. They were displayed and reconstructed using a computer program (see Gobet & Simon, 1996d; 1998 for details). Each position was displayed for 8 seconds, and the inter-position latency was 2 seconds, during which the screen was blank.

The random selection of the positions meant that some positions were taken in the middle of an exchange or of some other tactical complications. Such positions are normally not used in chess research, because players find them distracting. Therefore, the positions that P received were on average less typical that the positions normally used in chess research, which made them somewhat harder to memorize. S increasingly got used to the presence of these atypical positions.

The experimental sessions took normally place from Monday through Friday, between 9 am and 10 am, in a quiet room. Each session had two parts. First, there was the presentation of two warm-up positions. Then, after a short break, there was

[3] The skill level of competitive chess players is measured using the international Elo rating scale, which has a theoretical mean of 1500 points and standard deviation of 200 points.

the multiple-position task proper. The progression rules for this second task were as follows: (a) the minimum number of positions was four; (b) if more than one position in the previous session fell below 60% correct recall, then the number of positions was decreased by one; else, this number was increased by one.

P had the freedom to pace the two parts of a session as he wished. The typical behavior was as follows. At the very beginning of the experiment, P would select the cue list (see below) matching the current number of positions that he would receive on that day, and concentrate on it going through the names sub-vocally. He would then do the two-position warm-up (the cue list was not used for these two positions), then concentrate again on the cue-list, and finally perform the multiple-board task.

Overall Results

Inspired by Chase and Ericsson's (1981) research on the digit-span, P created a mnemonic system aimed at facilitating LTM encoding of chess positions. For sequences of four positions and more, he attempted to associate each position with the corresponding element in a pre-learned list containing the chess world champions in historical order (see Table 2 and Figure 1). The list of world champions was used so that P could fairly easily create meaningful associations between the position currently being displayed and the corresponding name in the list. Associations were mediated by verbal labels. When attempting to recall the positions, each name in the list would serve as a retrieval cue. After about 10 weeks, the names were abbreviated to their first syllables so that they could be pronounced more rapidly sub-vocally.

-- Insert Table 2 --

-- Insert Figure 1 --

The following examples, taken from Gobet and Simon (1996d), illustrates how the cue list was used. The examples go from rich associations (cases 1 and 2, where the type of position is recognized and an association made with the name in the cue-list) to poor associations (in case 3, a verbal label coding only two pieces is associated, and, in case 4, there is failure to associate an otherwise useful label with the current name in the cue list).

(1) Position #5. Name on the list: Euwe.

"A Panov attack. Black has a strong Knight on d5, typical for Euwe's play."

(2) Position #6. Name on the list: Botvinnik.

"A Grünfeld defense, as in the match Karpov-Kasparov, Seville. Botvinnik used to play the Grünfeld."

(3) Position #1. Name on the list: Steinitz.

"White has the Bishop pair. Steinitz liked the Bishop pair."

(4) Position #2. Name on the list: Lasker.

"A Maroczy without g6."

Figure 2 shows the number of pieces recalled as a function of session number, and Figure 3 shows the number of positions attempted also as a function of session number. Both Figures suggest that there is first a period of slow improvement, roughly until session 80, followed by a long plateau with little, if any, improvement. This impression is confirmed by statistical analysis. A linear regression analysis shows that, until session 80, session number is a statistically significant predictor of the number of pieces correct (pieces = 57.4 + 0.54 × session_number; $r^2 = .38$, $p < .001$) and boards attempted (board_attempted = 4.1 + 0.024 × session_number, $r^2 = .26$, $p < .001$). Thus, P gains about half a piece per session and it takes about forty sessions to increase the number of attempted boards by one. From session 81 to 213, there is no linear relationship (pieces correct by session: $r^2 = .005$, *ns*; boards attempted by session: $r^2 = .001$, *ns*). Note also that P managed to recall 150 pieces or more in 12 cases only, and attempted 10 positions in 3 cases only.

-- Insert Figure 2 --

-- Insert Figure 3 --

Specialization Effects in Expertise: The Question

If, as argued by De Groot (1965), Chase and Simon (1973b), and others, expertise relies primarily on perceptual information encoded either as knowledge complexes or chunks, then it follows rather directly that it should be difficult to transfer one's expertise in a given domain to another one. The perceptual chunks acquired over years of practice and study, which enable fluid behavior in the first domain with rapid recognition of the key features, simply will not be useful in other domains. We know at least since Thorndike and Woodworth (1901) that transfer between domains is difficult, and a recent meta-analysis shows that chess, the domain addressed by the chapter, is no exception (Gobet & Campitelli, 2006). But what about transfer between several within-domain specializations – for example neurology and pediatrics, or expertise in Chinese foreign policy and expertise in India's parliamentary system?

Theories emphasizing specific perceptual knowledge, including chunk-based theories, still make the prediction for domain-specificity, that is, that experts with one specialization should be better in that domain than experts with other specializations. However, this view is not shared by some theorists. For example, in his discussion of chess expertise, Holding (1985) claimed that the main components of skill is not pattern recognition, but forward search and general analytical reasoning abilities. So, Holding (1985, pp. 249-250) argued that “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.” More recently, Linhares (2005) has criticized the emphasis on chunks and templates in problem solving and argued that experts used “abstract roles” deriving not on surface, perceptual information, but on the deep meaning of chess positions. (How this deep meaning is accessed, however, is not spelled out in detail.)

Another criticism of the theoretical role of specialized chunks has been put forward by Ericsson and colleagues, who argued that superior performance can be achieved with relative small amounts of practice. SF and DD practiced just several hours a week during about 2 and 3 years, respectively, but had a better memory for digits than professional memory experts having practiced for more than 20 years of experience (Chase & Ericsson, 1982). Similarly, a novice trained by Ericsson and Harris (1990) was able, after about 50 hours of practice, to recall unfamiliar chess positions to the level of experienced chess players having spent thousands of hours practicing and studying chess. Based on these studies, Ericsson and Kintsch (2000) concluded that “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). However, their argument fails to consider that SF and DD used a powerful coding strategy and brought to the digit-span task extensive knowledge of numbers and dates, and the novice trained by Ericsson and Harris acquired only one component of the knowledge necessary for chess expertise, namely perceptual chunks. That this last result is consistent with chunking and template theories has been shown by Gobet and Jackson (2002), who replicated Ericsson and Harris's study and showed that simulations with CHREST (see below) could account for the results very well.

Specialization Effects: Method of Analysis and Results

Chess players specialize in the type of openings they use – it is simply impossible to know enough about all openings to be able to play them at a competitive level. In addition, it is important to find openings that match one's style of play, and an important role of coaches is to help players select such openings. Each player thus develops an "opening repertoire," for both White and Black. To avoid being too predictable, and also to anticipate the possibility that an opening would be in crisis because of the discovery of some new way to handle it, chess masters typically build some variability in their repertoire. For example, they could have two different replies against White's 1.e2-e4.

P was no exception to this, and had chosen a subset of openings that he studied in great detail and that he regularly used in competitive games, although there were a few changes over the years. He was able to identify apparently minute differences in the positions arising from the openings in which he specialized, and these differences could make the difference between a win and a loss. With the openings that did not belong to his repertoire, he would know the general ideas and plans, but would not be able to discriminate positions in the same way. It is then possible to use the data of the training experiment presented above to directly test the hypothesis that experts encode specific perceptual patterns. If this hypothesis is correct, P should on average obtain better recall performance with positions coming from his pet openings, or with positions resembling those positions.

In order to carry out the analysis, the stimulus positions were coded as (a) belonging to the type of openings P used to play, (b) belonging to openings P never played, and (c) positions that could not be classified (e.g., because too few pieces were left on the chessboard). Note that all stimulus positions were taken after White's twentieth move. This coding was carried out by P himself, being the best expert of the kind of openings he had used during his chess career. The following analyses will be on the first 600 hundred positions used in the training experiment.

-- Insert Figure 4 --

As Figure 4 shows, P better remembered the type of positions he used to play than the other types of positions, with the exception of the cases where he attempted 9 and 10 boards. The differences are statistically significant both in the played vs. non-played comparison, $t(7) = 6.52$, $p < .001$ (one-tailed), and in the played vs. unclear comparison, $t(7) = 3.26$, $p < .01$ (one-tailed). In general, the positions coming from openings he did not play had the lowest recall. These data thus add support to the hypothesis that the encoding of specific perceptual information plays an important role in chess expertise and that restricting task practice to a particular subset of the task environment constrains transfer of skill, hence performance, outside of the selected practice environment. At the same time, it is important to note that P did not do that badly (62% correct) with positions that did not belong to his repertoire, as compared to 75% correct with positions that did. It is reasonable to suspect that knowledge – either perceptual chunks or general knowledge – he had acquired during his training and practice (for example, by replaying games of famous players) allowed him to encode at least part of most unfamiliar positions and thus achieve this level of recall.

Of course, the chunking and template theories, which account for participants' performance in other studies, predicted this result. It is thus important to check whether computational models based on these theories can simulate it. In the next section, I address this question by reporting computer simulations with CHREST.

Specialization in Chess: Computer Simulations

The CHREST Model

Converging evidence for current claims about chess specialization come from computer simulations that used CHREST (Chunk Hierarchy and REtrieval STructures), a model implementing important aspects of the template theory (Gobet & Lane, 2005; Gobet et al., 2001; Gobet & Simon, 2000). The first version of CHREST was developed in the early nineties to provide a unified model of perception and memory in chess. In particular, the aim was to put together the key insights of chunking theory and previous models of chess perception (Simon & Barenfeld, 1969) and memory (Simon & Gilmartin, 1973). The development of the model was also informed by the collection of new experimental data (Gobet & Simon, 1996a, 1996b, 1996c, 1996d) and theoretical reflections that led to the template theory. The later versions of CHREST were more ambitious with respect to their scope, first covering other domains of expertise and later addressing cognition more generally. To date, the CHREST architecture has been used to simulate phenomena in a number of domains including concept formation (Lane & Gobet, 2007), problem solving in physics (Lane, Cheng, & Gobet, 2000) and awele (an African game; Gobet, 2009), children's acquisition of vocabulary (Jones, Gobet, & Pine, 2007), and children's acquisition of syntax (Freudenthal, Pine, Aguado-Orea, & Gobet, 2007). CHREST's ability to simulate these phenomena illustrates the explanatory power of the chunking hypothesis, something Bill Chase was undoubtedly aware of. The following paragraphs provide a rather basic introduction to CHREST. A fuller description is provided in Gobet and Chassy (2009) and the computer code can be found at www.CHREST.info.

As the simulations will focus on chess, a domain where the processing of verbal information plays only a limited role (see Gobet, de Voogt, & Retschitzki, 2004, for details), description of CHREST will focus on the visual aspects of the model. CHREST comprises four main components: an LTM where chunks are stored, a visual STM, a simulated eye, and a mind's eye, where visuo-spatial information can be manipulated. Figure 5 illustrates these four components. The simulated eye moves around the chessboard, fixating squares, and the information within the visual field is sent to a discrimination network in LTM which mediates recognition of a chunk. A pointer to this chunk is then placed in STM, and the information is also unpacked in the "pictorial short-term memory," another name for the mind's eye. This sequence of operations is then repeated.

--- Insert Figure 5 ---

With respect to LTM, the interest focuses on the creation and use of chunks. Chunks, together with the links that connect them, form a discrimination net. The links have tests, which are applied to check features of the external stimuli. This component of the model borrows several mechanisms from Feigenbaum and Simon's (1984) EPAM model. As noted above, chunks that are recognized often evolve into more complex data structures. The core of templates is similar to the information stored in chunks, and this information is stable. Templates also have slots, which make it possible for the value of variables to be encoded rapidly. Based on computer simulations, it is assumed that it takes 250 ms to instantiate a template slot. In the current simulations, slots can encode information about piece location, piece type, or chunks. Slots are automatically created where the links irradiating from a given chunk show enough variation on a given type of information (e.g., information about squares, type of pieces, or groups of pieces; see Figure 6). Template slots are created when the number of nodes that are linked to a given node and share identical information is greater than a parameter, arbitrarily set to 3 in the simulations. A further constraint is that a chunk should contain at least 5 elements. Chunks and templates can be linked to other chunks and templates, as well as other information stored in LTM, most notably moves and sequences of moves.

-- Insert Figure 6 --

Visual STM has a capacity of three chunk pointers. When STM is already full and a new chunk is recognized, the oldest pointer leaves STM. There is an exception to this rule with the largest chunk: its pointer leaves STM only when a larger chunk is recognized. This idea of pointers has often been criticized on the grounds that it is too close to computer programming and lacks biological plausibility, but this criticism is misguided. A neurally plausible implementation for such pointers has been proposed by Ruchkin, Grafman, Cameron, and Berndt (2003). These authors suggest that STM neurons in the prefrontal cortex fire in synchrony with neurons located in posterior areas of the cortex. This mechanism has the advantage of explaining why STM has a limited capacity; it is due to the limited number of distinct frequencies available for synchrony, which limits the number of pointers available.

Eye movements that serially scan a position are directed from a combination of heuristics (e.g., heeding a square on a part of the board that has not been visited by the eye yet) and acquired knowledge, which is mediated by the structure of the discrimination net. Specifically, the model uses the largest chunk recognized so far to direct eye movements (see De Groot & Gobet, 1996, for details).

The final component of CHREST, the mind's eye, clearly shows the influence of Chase and Simon's (1973a) paper *The mind's eye in chess* on the model. The mind's eye is similar to Kosslyn's (1994) visual buffer and Baddeley's (1986) visuo-spatial sketchpad, and stores perceptual structures, both from memory stores and external inputs, for a short amount of time. Its content is encoded as a network of nodes and links that can be manipulated by visuo-spatial mental operations. It is worth pointing out that the information encoded in the mind's eye is much more abstract than the information impinging the retina. Unless it is refreshed, information in the mind's eye suffers from rapid decay, within around 250 ms (Kosslyn, 1994). The processes taking place in the mind's eye (e.g., the time to move a piece mentally) are assumed to be carried out serially; Kosslyn, Cave, Provost, and Von Gierke (1988) provide data supporting this assumption. Finally, CHREST includes mechanisms linking LTM, STM, and the mind's eye. When a chunk is elicited, either by external or internal information, a pointer to it is placed in STM. Concurrently, the visuo-spatial information referred to in LTM by this pointer is unpacked in the mind's eye. As information in the mind's eye fades rapidly, it needs to be refreshed regularly. Waters and Gobet (2008) provide more details on how CHREST's mind's eye works, and present an experiment testing some of the assumptions discussed in this paragraph.

Four characteristic features of CHREST might be highlighted. First, there an emphasis of cognitive limitations – for example limitations in memory capacity (visual STM can hold only three items) and learning rates (it takes about eight seconds to create a new chunk). This emphasis can be traced back to Simon's work on bounded rationality (e.g., Simon, 1969). Second, CHREST is a self-organizing, dynamical system. The structure of its discrimination net cannot be predicted on general considerations alone, and depends both on internal mechanisms and on the statistical structure of the environment. Third, and closely related to the previous point, learning and thus the quality of the simulations depend on realistic input capturing the detail of the structure of the environment. Finally, time parameters are used for each cognitive process (e.g., 8 seconds to create a new chunk, or 50 ms to encode a chunk in STM), which makes it possible to derive detailed quantitative predictions.

Simulation Methods

The simulations aim to establish whether CHREST's structures and mechanisms are sufficient for reproducing the specialization effect found in P's training experiment. This investigation uses the same model as that used in previous simulations (Gobet & Simon, 2000; Gobet & Waters, 2003; Waters & Gobet, 2008), including the same mechanisms for template creation and use. The only novel addition is that the model also uses a retrieval structure, which is similar to the list of world champions used by P for the recall of four and more positions (see Table 2 and Figure 1). For the recall of two positions (warming-up), a retrieval structure with two items (encoding the positions as "first" and "second") is postulated. The probability of storing a chunk or template into the retrieval structure was set at 1 with 2 positions and .84 with 4 and more positions.

During the learning phase, the program incrementally acquired chunks and templates by scanning a large database of positions (about 50,000) taken from master-level games. With CHREST, skill differences are simulated by growing nets of various sizes. For the purpose of these simulations, the learning set consisted of a mixture of positions related to P's opening repertoire (about 90%), positions taken from P's games (about 200, shown 10 times, 5%), and positions randomly selected from databases (5%). These numbers were supposed to reflect the kind of positions P was familiar with: the most familiar were the positions from games he had played, followed by positions he was likely to have studied, and then finally positions from openings he did not play. The network used in the simulations had ~300,000 chunks and ~20,000 templates.

Results

As shown in the lower panel of Figure 4, the model does a good job at capturing the difference between positions belonging to P's repertoire and other positions. Like P, the model recalls the openings from his repertoire better than openings he did not play, $t(7) = 5.46$, $p < .001$ (one-tailed), and than openings difficult to classify, $t(7) = 4.97$, $p < .002$ (one-tailed). The model replicates the specialization effect because it can find more chunks and in particular more templates in the positions that were predominantly used for training (i.e. positions belonging to P's repertoire). However, it should be pointed out that the model performs rather consistently about 10% below the level reached by the player. This probably could be corrected by using a network with more chunks and templates for the simulations.

In summary, the simulations showed that CHREST's mechanisms for creating and using chunks and templates lead to knowledge that is consistent with the specialization effect observed in the single-case experiment. They also show that part of this knowledge can be recruited to recall positions that do not belong to the domain of specialization of the model, albeit with weaker recall performance.

Specialization in Chess: French Connection and Sicilian Moves

The case study presented earlier offered relatively clear results, and the computer simulations provided a mechanism for them. But do the results generalize to a larger sample? Do they apply to other tasks such as problem solving, where one has to find the best move in a position? A recent study (Bilalić, McLeod, & Gobet, 2009) addressed these very questions.

In this study, we combined two research paradigms typical of expertise research. The first paradigm, the most common and that used for example in Chase and Simon's study, consists of comparing individuals of different skill levels. The second paradigm consists of comparing groups of individuals with the same skill level, but with different fields of specialization. This second paradigm has the advantage that factors such as amount of experience and general skill level are controlled for; what differs between the two groups is the subset of domain knowledge individuals specialized in. Researchers have used the specialization paradigm for studying expertise in medicine (Rikers et al., 2002), political science (Chiesi, Spilich, & Voss, 1979), and the design of experiments (Schunn & Anderson, 1999). In general, the results indicate that, when specific domain knowledge is not available, experts fall back on weaker methods. These results have sometimes been taken as contradicting theories of expertise such as the chunking theory, which emphasize the role of domain-specific patterns and methods (e.g., Schunn & Anderson, 1999). There is no doubt that experts use general problem-solving methods in addition to domain-specific methods. The interesting question, however, is how expert performance changes when only domain-general methods are used.

The study presented in Bilalić et al. (2009) is based on one of the experiments that Merim Bilalić carried out for his PhD thesis. Bilalić's elegant design addressed some of the weaknesses of previous studies using the specialization paradigm. For example, many studies used one type of problem, which has the disadvantage that it is from the field of one group of experts but not of the other. Bilalić et al. (2009) presented two kinds of problems, which mapped onto the specialization of each group. Also, few studies used neutral problems (i.e. problems unfamiliar to both groups), which seem necessary to tease apart the effect of knowledge. Finally, in many studies, the participants' level of expertise, and therefore the effect of expert specialization, was difficult to quantify.

Chess offered a great domain for addressing these issues. The Elo rating makes it possible to measure skill precisely; chess players love trying to solve chess problems; and the fact that players specialize in different openings, as explained in the case study above, makes it easy to find players of the same skill level (as measured by the Elo rating) but with different knowledge for a specific opening.

-- Figure 7 --

Method

Thus, Bilalić's idea was to compare performance of players within and outside their domain of specialization. He used three types of middle game positions: positions from the Winawer variation of the French defense (see Figure 7, left), positions from the Najdorf variation of the Sicilian defense (see Figure 7, right), and neutral positions. The neutral problems came from middle game positions difficult to classify with respect to the opening they came from, and were used to measure more general memory and problem-solving abilities. There were two tasks. In the memory recall task, positions were presented for 5 seconds. The stimuli consisted of 12 positions (4 positions for each of the French, Sicilian, and neutral conditions) that were presented in random order. In the problem-solving task, players were required to think aloud while trying to find the best move. They were not allowed to move the pieces, and there was no time limit. There were two problems for each of the three kinds of positions, and positions were presented in a random order. It took on average one hour to the participants to go through all six problems.

The participants were 8 Candidate Masters (with an average of 2140 Elo), 8 Masters (2300 Elo), and 8 International Masters/Grandmasters (2490 Elo). Half the players specialized in the Winawer variation of the French defense, and the other half in the Najdorf variation of the Sicilian defense.

Results

Figure 8 shows the results for the memory task. In line with previous experiments, there was a reliable skill effect, with better players obtaining better scores than less skilled players, $F(2, 18) = 19.88$, $p < .01$, $\eta_p^2 = .69$. The key result was that players were better at recalling positions from their opening repertoire than positions they did not play. Thus, French players recalled French positions better, and Sicilian players recalled Sicilian positions better, yielding a significant interaction player type × position type: $F(1, 156) = 46.96$, $p < .01$, $\eta_p^2 = .23$. The effect of familiarity is strong enough to take precedence over skill. With the Sicilian positions, the performance of Sicilian candidate masters was close to the performance of French masters, and the performance of Sicilian masters was close to that of French grandmasters/international masters. The same pattern of results was observed with French positions, where French candidate masters and masters obtained comparable results to those obtained by Sicilian masters and grandmasters/international masters, respectively. Thus, the magnitude of the specialization effect on chess memory is about one standard deviation in skill (see footnote 3). As both groups memorized the neutral middle-game positions equally well, one can be confident that the two groups were of similar strength.

-- Insert Figure 8 --

-- Insert Figure 9 --

Move quality in the problem-solving task was computed as the difference between the assessment of the best move in the position, as estimated by Fritz 8 (a very strong computer chess program) and the chosen move. Note that Fritz 8 evaluates moves using pawn units. Thus, to give an example, a score of 0.5 means that the selected move was inferior by half a pawn to the best move in the position.

The problem-solving task yielded very similar results to the memory task (see Figure 9). There was a reliable skill effect, better players choosing better moves, $F(2, 18) = 7.35$, $p < .01$, $\eta_p^2 = .45$. Crucially, the interaction player type × position type was statistically significant. Players faced with positions from within their opening specialization found better moves than those faced with positions outside it, $F(1, 64) = 13.87$, $p < .01$, $\eta_p^2 = .18$. Moreover, the magnitude of the specialization effect – about one standard deviation in skill level – was comparable to that found in the memory task. With the French positions, the French candidate masters and masters found slightly better moves than the Sicilian masters and grandmasters/international masters, respectively. (It is unknown why the Sicilian Masters performed so poorly in the French positions.) The effect was even stronger with the Sicilian positions.

In summary, players outside their domain of specialization perform about 200 Elo points (one standard deviation in skill) below their level with familiar positions, both with a memory and problem-solving task. Thus, a grandmaster (≈2600 points) performs at the level of an international master (≈2400 points). Or, to put it another way, a player ranked #186 in the world would perform roughly like a player ranked #2837 in the world.

Specialization in Science

Rather unexpectedly, the results I have presented here have implications for policy, more precisely for science policy in the context of peer review. The quality of the research undertaken by all UK universities is evaluated about every 5 years, in a mammoth data collection exercise previously called the Research Assessment Exercise (RAE; www.rae.ac.uk), and renamed in 2007 as the Research Excellence Framework (REF; www.hefce.ac.uk/research/ref). The rankings obtained are used to allocate a subset of research funding (the equivalent of about $2.5 billion a year). The RAE is thus no laughing matter, and universities spend a considerable amount of time and money to optimize the information they return.

The key evaluation criterion is the quality of research papers produced by a department. Each academic discipline is judged by a panel of experts (13 for psychology) coming from the same discipline. The panel members read the best four papers of each academic, and rate them.[4] The scale used has changed over the years, and included five levels in the last RAE, conducted in 2008. And this is here that the specialization effect crops in, I believe.

Panels have to evaluate a wide range of papers. For example, in psychology, papers would cover subfields as varied as clinical psychology, social psychology, cognitive neuroscience, and cross-cultural psychology, to mention just a few. And, of course, each sub-discipline will carry out research that has little in common with other sub-disciplines. Think for example, within cognitive psychology, of the research carried out in perception with fMRI and in problem solving with protocol analysis.

In these circumstances, is the judgment of the panel experts reliable? We simply do not know, as only the aggregated results are published, and the evaluation of individual papers is destroyed. However, the Bilalić et al.'s (2009) data clearly suggest that the RAE approach is not reliable: Top experts are likely to behave like average experts when outside their domain of specialization. They still will be experts, but their level will be far from the level of expertise for which they have been selected in the first place.

One could object that it is inappropriate to generalize from chess to the evaluation of scientific papers. I would argue that, if anything, the situation is probably worse with the RAE than with chess. Evaluating the quality of scientific papers is a subjective process, prone to noise. In chess, a checkmate in three moves does not leave much space for discussion. In science, whether an idea is good or bad,

[4] This of course imposes a considerable burden on the panel members – the time spent evaluating papers is not spent on their own research. For an interesting and early discussion of the cost of reviewing, as well as other related questions such as the efficiency of the peer-review system at the NSF, see Klahr (1985).

or an experimental result important or not, can lead to endless deliberations – just think how even experts in the same subfield disagree when it comes to review manuscripts or grant applications. Finally, the distance between specializations in science seems larger than between specializations in chess. Chess openings might differ, but the strategies and tactics for handling the middle game and the endgame tend to converge, irrespectively of the first moves. In science, such convergence is less likely, in great part due to differences in methods – evaluating research using brain imaging is miles apart from evaluating research on Freudian psychotherapy.

If the results of Bilalić et al.'s (2009) experiments generalize to the type of evaluation done in the RAE, then there is room for doubt that the assessors were properly qualified for their task. In other experiments carried out within the framework of Bilalić's PhD research, it was found that, as expertise diminishes, the likelihood of being victim to biases in thinking increases (Bilalić, McLeod, & Gobet 2008a, 2008b). Having papers evaluated by two experts, as was usually done in the RAE, might reduce this effect, but it is not clear that two experts outside their domain of special expertise can reach the level of an expert from within the domain.

Adaptability and transfer

In order to survive, organisms must adapt to their environment. As noted by Simon (1969), among others, one of the means to do so is to learn. Thus, the chunking and template mechanisms I have discussed in this chapter contribute in important ways to humans' adaptation. Their strengths are that they capture the statistical regularities of the environment whilst providing at the same time enough flexibility to insure adaptability to unexpected events. But what happens when the environment changes? What happens when experts face a new domain?

One must distinguish between two types of changes. In the first case, the new environment still overlaps to a large extent with the old one. This is the case that was considered in this chapter. Within the same domain, experts in one sub-domain (e.g., chess players specializing in the Sicilian defense) faced problems from a different sub-domain (e.g., positions from the French defense). A decrease in performance was found, but not a drastic one. Obviously, players could use general heuristics or other types of abstract knowledge to cope with the less familiar situations. They could also use knowledge coded as chunks and templates, as suggested by the CHREST simulations. While optimized for a specific environment, these knowledge structures are flexible enough that at least some of them can be used in a different, but related environment.

In the second case, the new environment has little to do with the old one. This is of course the classic question of transfer studied in psychology and education for more than one century. Here, researchers, starting with Thorndike and Woodworth (1901), have been more pessimistic about the possibility of using knowledge structures in the new domain. For example, while DD was able to memorize more than one hundred digits, this skill did not transfer to the memory for words (Chase & Ericsson, 1982). Similarly, contrary to a widely held opinion in the chess community, skill in chess does not transfer to other domains such as English and mathematics (Gobet & Campitelli, 2006). Rare exceptions to this pattern of results are offered with transfer from violent action video games (Green, Li & Bavelier, 2009; see chapter ? of this book) and by the transfer of highly abstract kinds of knowledge, such as mathematical techniques.

However, contrasting with these rather depressing conclusions, “real-world” experts offer some spectacular counter-examples. Eric Heiden is widely considered as the best speed skater of all times. Among other feats, he won all five races in the 1980 Olympics games in Lake Placid, both in sprint and endurance, establishing four Olympic records and one world record. In his second career, he was a professional road racing cyclist, winning the USA championship in 1985. In his third career, he specialized in orthopedic surgery after obtaining an MD. Simen Agdestein carried out two international careers in parallel: the first as a chess grandmaster (he was seven times Norwegian champion) and the second as a football player (he played eight times for the Norwegian national team). Later, he became a successful chess coach, having notably advised current chess number 1 Magnus Carlsen. And, of course, there is Arnold Schwarzenegger. First, one of the most successful body builders ever (five Mr. Universe titles and seven Mr. Olympia titles); then a successful businessman in bricklaying business and real-estate investing; then one of the most famous actors in the world; and finally a successful politician (Governor of California).

How can we explain this discrepancy between academic research and actual expertise? Do these individuals use better heuristics? Do they somehow acquire chunks and templates that are more generalizable? Have they developed a particular efficacious type of opportunism, optimizing the reuse of assets acquired in a previous career? Is it "only" a question of talent – e.g. better motivation, will power and stamina? Surprisingly, little research has addressed these issues.

Summary and Conclusions

As this chapter has made it clear, my research has been greatly influenced by Bill Chase, although I never had the pleasure to meet him. His influence includes the study of specialized knowledge in expertise, the general idea of chunking as a key theoretical mechanism of human cognition, and the use of single-subject designs. The overall message of this chapter is that specialization plays a powerful role in determining expert performance. This was first shown in a memory task with a single expert. Computer simulations with CHREST replicating the specialization effect contributed converging evidence. An additional study with a larger sample showed that the result could be generalized beyond memory tasks to problem solving. In general, the results support the link between memory of specific experiences and problem-solving ability; thus, problem-solving strategies that are context-independent and general-purpose are not sufficient to make somebody an expert. This is consistent with what Bill Chase claimed about domain-specificity in the skilled memory effect (Chase, 1986).

References

Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.

Bilalić, M., Mcleod, P., & Gobet , F. (2008a). Inflexibility of experts - Reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive Psychology, 56*, 73-102.

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

Bilalić, M., McLeod, P., & Gobet , F. (2009). Specialization effect and its influence on memory and problem solving in expert chess players. *Cognitive Science, 33*, 1117-1143.

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

Campitelli, G., Gobet, F., Williams, G., & Parker, A. (2007). Integration of perceptual input and visual imagery in chess players: Evidence from eye movements. *Swiss Journal of Psychology, 66*, 201-213.

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

Chase, W. G. (1986). Visual information processing. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance (Vol. III, Cognitive processes and performance)* (pp. 1-73). New York: Wiley.

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

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. *The Psychology of Learning and Motivation, 16*, 1-58.

Chase, W. G., & Simon, H. A. (1973a). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

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

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 257-273.

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. (1965). *Thought and choice in chess (first Dutch edition in 1946).* The Hague: Mouton Publishers.

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

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (First edition, 1885 ed.). New York: Dover.

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

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

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.

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8*, 305-336.

Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modelling the developmental patterning of finiteness marking in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science, 31*, 311-341.

Frey, P. W., & Adesman, P. (1976). Recall memory for visually presented chess positions. *Memory and Cognition, 4*, 541-547.

Gobet, F. (2000). Some shortcomings of long-term working memory. *British Journal of Psychology, 91*, 551-570.

Gobet, F. (2009). Using a cognitive architecture for addressing the question of cognitive universals in cross-cultural psychology: The example of awalé. *Journal of Cross-Cultural Psychology, 40*, 627-648.

Gobet, F., & Campitelli, G. (2006). Education and chess: A critical review. In T. Redman (Ed.), *Chess and education: Selected essays from the Koltanowski conference* (pp. 124-143). Dallas, TX: : Chess Program at the University of Texas at Dallas.

Gobet, F., & Chassy, P. (2009). Expertise and intuition: A tale of three theories. *Minds & Machines, 19*, 151-180.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four… or is it two? *Memory, 12*, 732-747.

Gobet, F., de Voogt, A. J., & Retschitzki, J. (2004). *Moves in mind: The psychology of board games*. Hove, UK: Psychology Press.

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

Gobet, F., & Lane, P. C. R. (2005). The CHREST architecture of cognition: Listening to empirical data. In D. Davis (Ed.), *Visions of mind: Architectures for cognition and affect* (pp. 204-224). Hershey, PA: IPS.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*, 236-243.

Gobet, F., & Ritter, F. E. (2000). Individual data analysis and Unified Theories of Cognition: A methodological proposal. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modelling* (pp. 150-157). Veenendaal, The Netherlands: Universal Press.

Gobet, F., & Simon, H. A. (1996a). Recall of random and distorted positions. Implications for the theory of expertise. *Memory & Cognition, 24*, 493-503.

Gobet, F., & Simon, H. A. (1996b). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review, 3*, 159-163.

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

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

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24*, 651-682.

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

Green, C. S., Li, R. J., & Bavelier, D. (2009). Perceptual learning during action video game playing. *Topics in Cognitive Science, 2*, 202-216.

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

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science, 10*, 853-873.

Klahr, D. (1985). Insiders, outsiders and efficiency in an NSF Panel. *American Psychologist, 40*, 148-154.

Kosslyn, S. M. (1994). *Images and brain: The resolution of the imagery debate*. Cambridge, MA: Bradford.

Kosslyn, S. M., Cave, C. B., Provost, D. A., & Von Gierke, S. M. (1988). Sequential Processes in Image Generation. *Cognitive Psychology, 20*, 319-343.

Lane, P. C. R., Cheng, P. C. H., & Gobet, F. (2000). CHREST+: Investigating how humans learn to solve problems using diagrams. *AISB Quarterly, 103*, 24-30.

Lane, P. C. R., & Gobet, F. (2007). Developing and evaluating cognitive architectures using behavioural tests. In *AAAI Workshop on Evaluating Architectures for Intelligence* (pp. 109-114). Hove, UK: Psychology press.

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

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

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

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

Richman, H. B., Gobet, F., Staszewski, J. J., & Simon, H. A. (1996). Perceptual and memory processes in the acquisition of expert performance: The EPAM model. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 167-187). Mahwah, NJ: Erlbaum.

Rikers, R. M. J. P., Schmidt, H. G., Boshuizen, H. P. A., Linssen, G. C. M., Wesseling, G., & Paas, F. G. W. C. (2002). The robustness of medical expertise: Clinical case processing by medical experts and subexperts. *American Journal of Psychology, 115*, 609-629.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences, 26*, 709-+.

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

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*, 250-264.

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Simon, H. A., & Barenfeld, M. (1969). Information processing analysis of perceptual processes in problem solving. *Psychological Review, 7*, 473-483.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist, 61*, 393-403.

Simon, H. A., & Gilmartin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5*, 29-46.

Staszewski, J. J. (1990). Exceptional memory: The influence of practice and knowledge on the development of elaborative encoding strategies. In F. E. Weinert & W. Schneider (Eds.), *Interactions among aptitudes, strategies, and knowledge in cognitive performance* (pp. 252-285). New York: Springer.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 9*, 374-382.

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

Table 1. Two complementary ways of summarizing data. In the traditional approach (second column), observables are summarized across subjects. In the Individual Data Modeling (IDM) approach (fourth column), theoretical parameters are first estimated for each subject using observables, and only then summarized across subjects. (After Gobet & Ritter, 2000.)

Subjects	Subjects' Observables (over tasks)				Estimated UTC Parameters
S_1	o_{11}, o_{12}, o_{13} ... o_{1t}	⇒	IDM	⇒	p_{11}, p_{12}, p_{13} ... p_{1n}
S_2	o_{21}, o_{22}, o_{23} ... o_{2t}	⇒	IDM	⇒	p_{21}, p_{22}, p_{23} ... p_{2n}
.					
.					
.					
.					
S_m	o_{m1}, o_{m2}, o_{m3} ... o_{mt}	⇒	IDM	⇒	p_{m1}, p_{m2}, p_{m3} ... p_{mn}
Summary values	O_1, O_2, O_3 ... O_t				$P_1,$ $P_2,$ $P_3,$ ... P_n

Table 2

List of the chess world champions used by P as a cue list (retrieval structure). After Gobet and Simon (1996d).

1.	Steinitz	Stein
2.	Lasker	Las
3.	Capablanca	Cap
4.	Alekhine	Al
5.	Euwe	Euw
6.	Botvinnik	Bot
7.	Smyslov	Smys
8.	Tal	Tal
9.	Petrossian	Pet
10.	Spasski	Spass
11.	Fischer	Fish
12.	Karpov	Kar
13.	Kasparov	Kas

Figure Captions

Figure 1. The method of *magistri*: The mnemonic system used by P consisted of combining a retrieval structure with templates.

Figure 2. Multiple-board experiment. Number of pieces correctly recalled as a function of session number.

Figure 3. Multiple-board experiment. Number of positions attempted as a function of session number.

Figure 4. Specialization effect in the multiple-board experiment: proportion correct as a function of number of positions attempted and type of positions. Top panel: human data. Bottom panel: computer simulations with CHREST.

Figure 5. Application of CHREST to the chess domain.

Figure 6. An example of template formation.

Figure 7. An example of a position coming from the French defense, Winawer variation (left) and from the Sicilian defense, Najdorf variation (right). In parentheses, the best move.

Figure 8. Specialization in a memory task: proportion correct as a function of the type of players, type of positions, and skill level. From Bilalić, McLeod, & Gobet (2009).

Figure 9. Specialization in a problem-solving task: deviation from the optimal move as a function of the type of players, type of positions, and skill level. Lower values indicate better moves. From Bilalić, McLeod, & Gobet (2009).

Figure 1

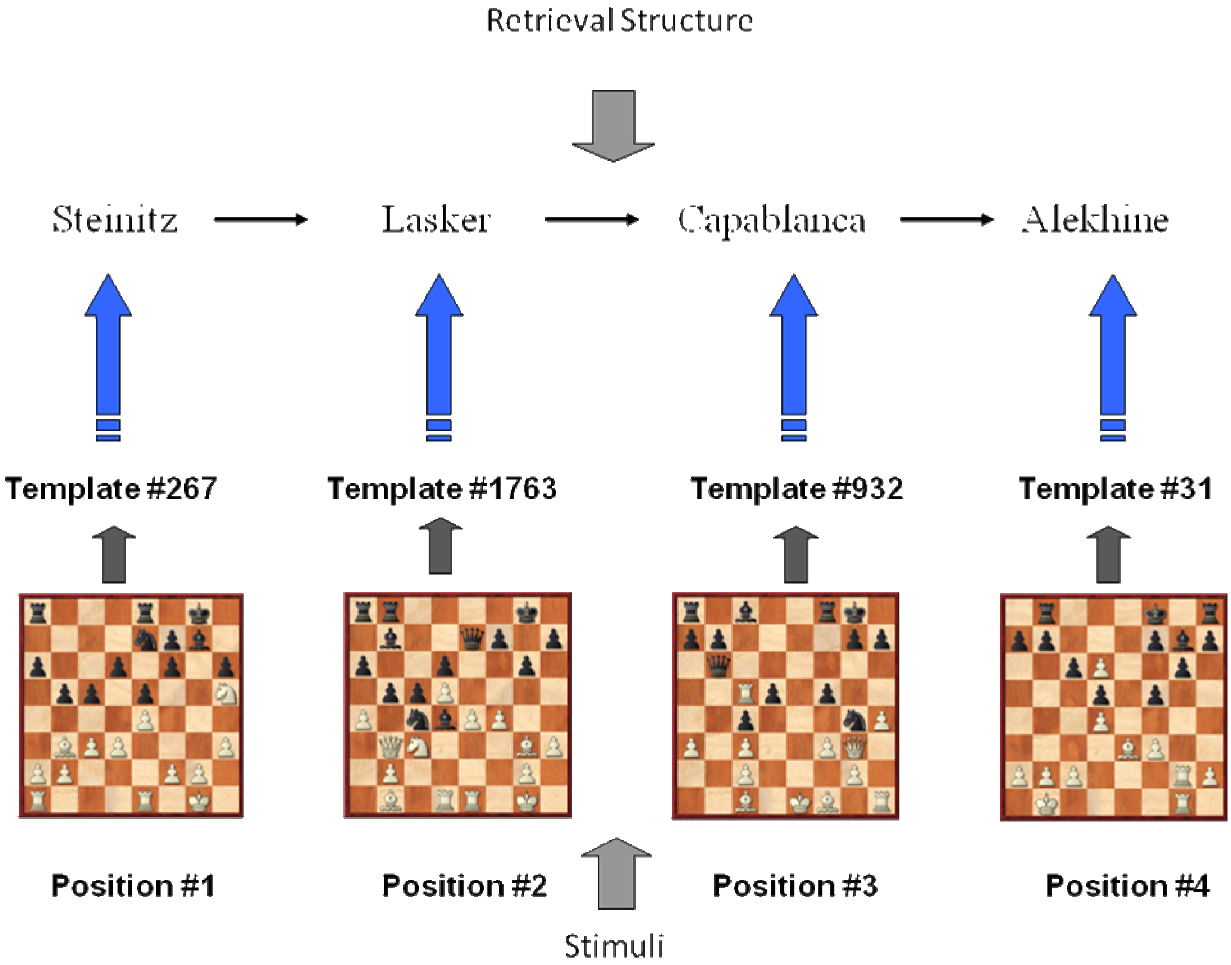

Retrieval Structure
Steinitz
Lasker
Capablanca
Alekhine
Template #267
Template #1763
Template #932
Template #31
Position #1
Position #2
Position #3
Position #4
Stimuli

Figure 2

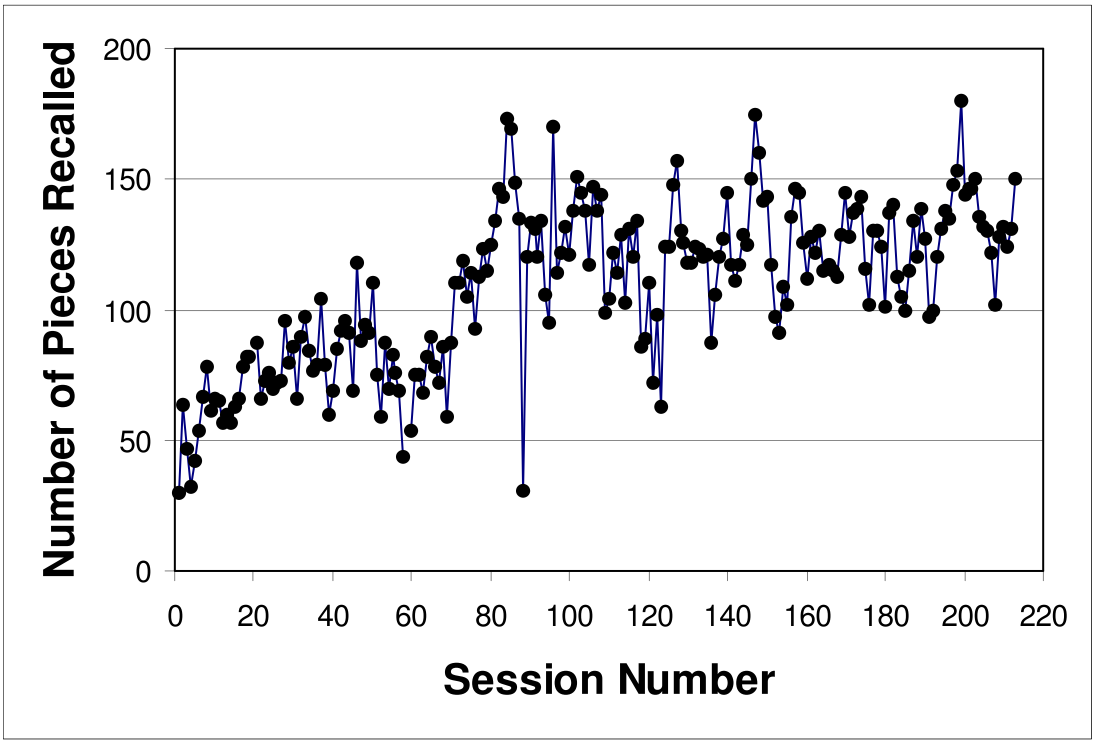

Number of Pieces Recalled
200
150
100
50
0
0
20
40
60
80
100
120
140
160
180
200
220
Session Number

Figure 3

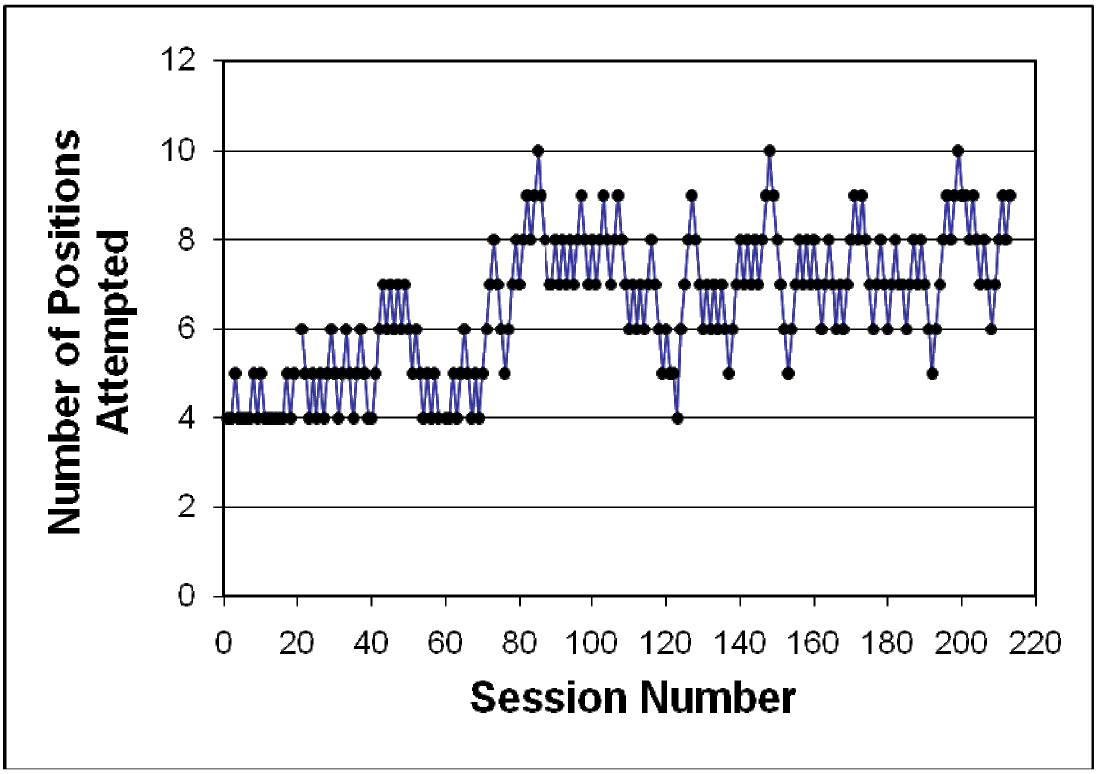
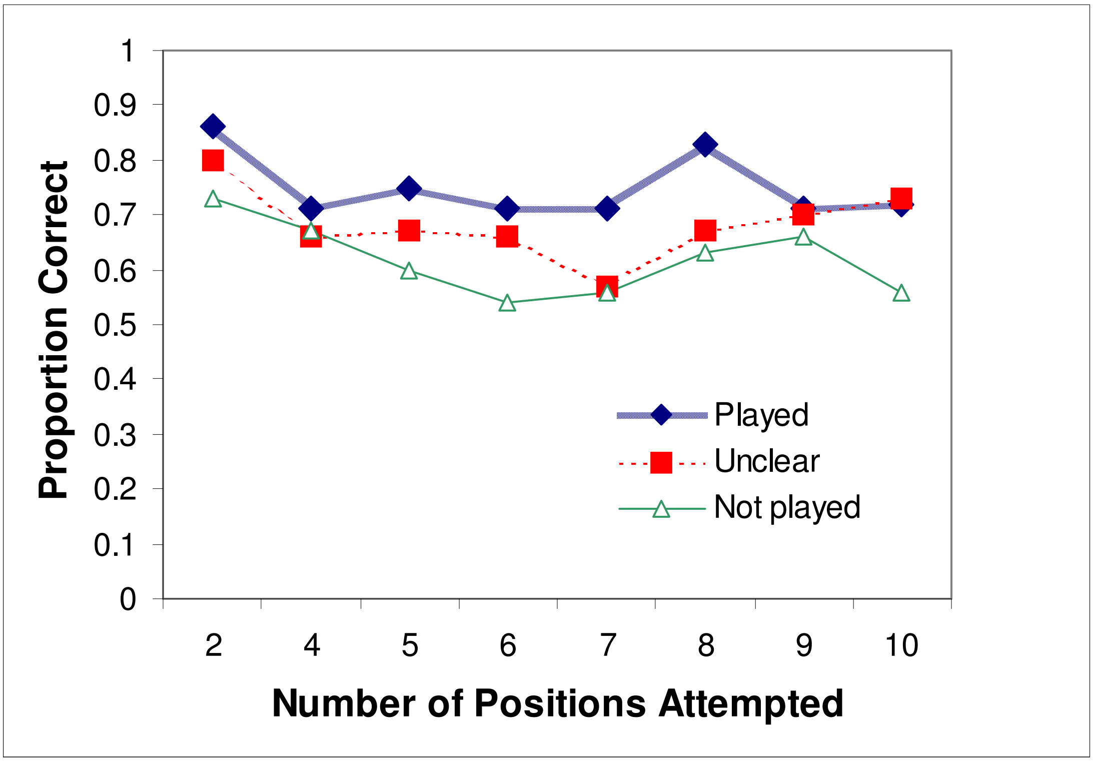

Number of Positions Attempted
12
10
8
6
4
2
0
0
20
40
60
80
100
120
140
160
180
200
220
Session Number

Figure 4

HUMAN PLAYER



CHREST

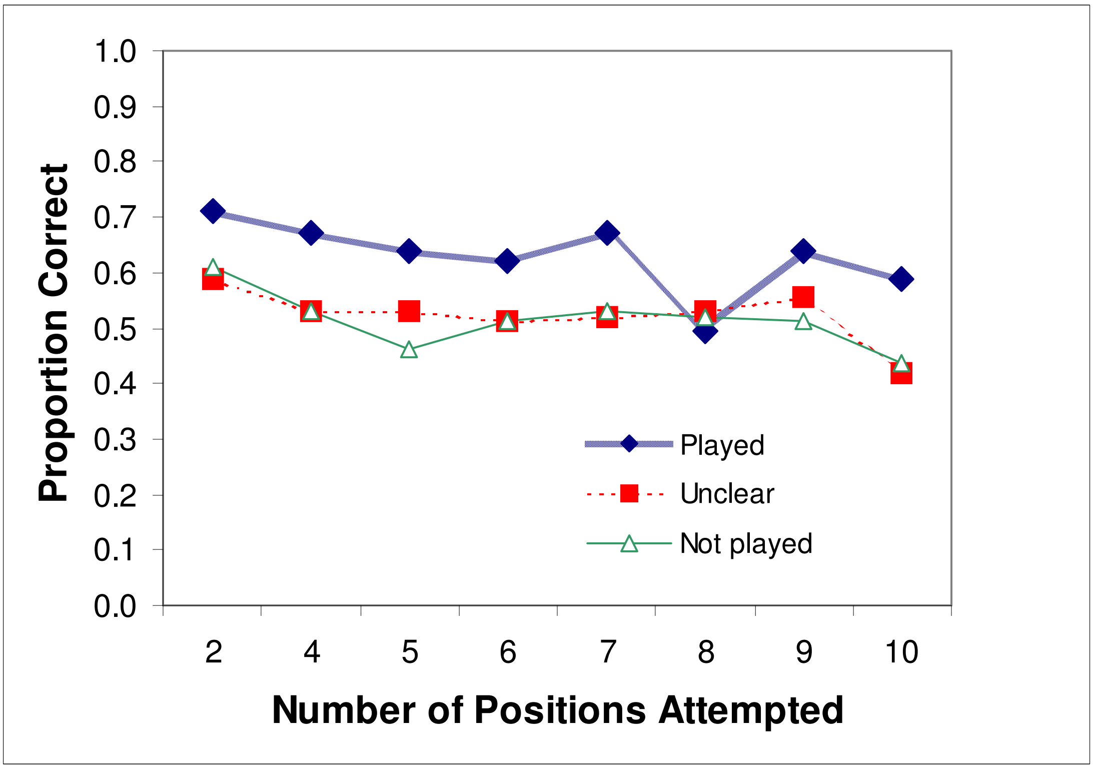


Figure 5

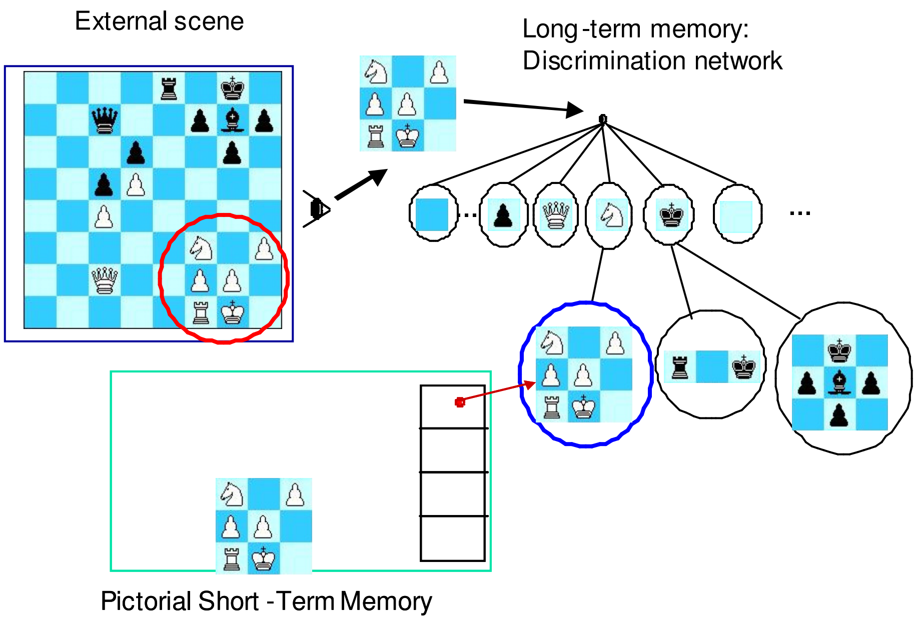

External scene
Long -term memory:
Discrimination network
...
...
Pictorial Short -Term Memory

Figure 6

Template formation

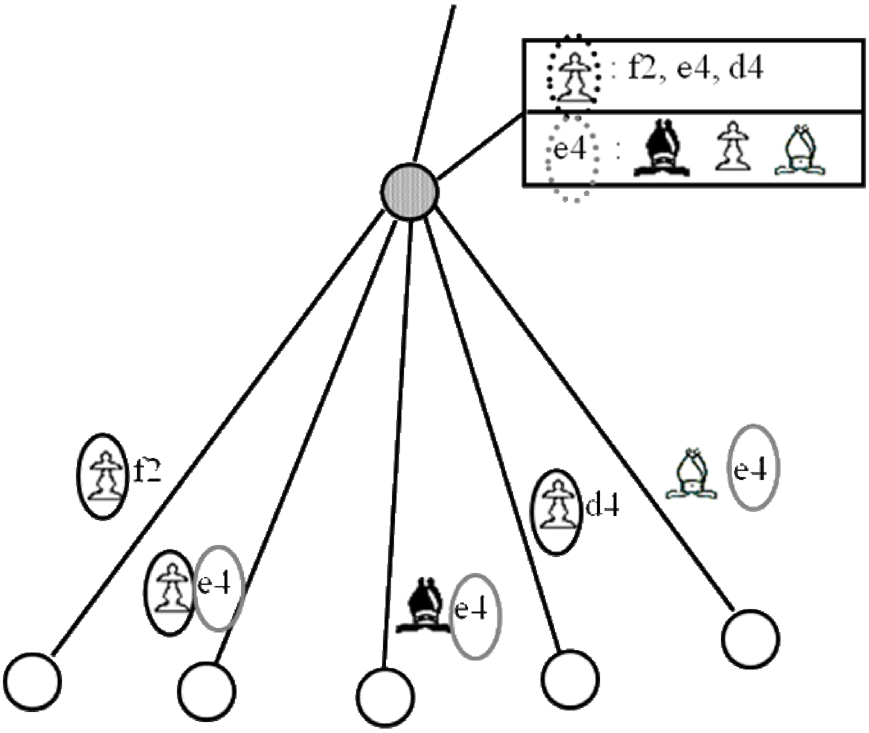
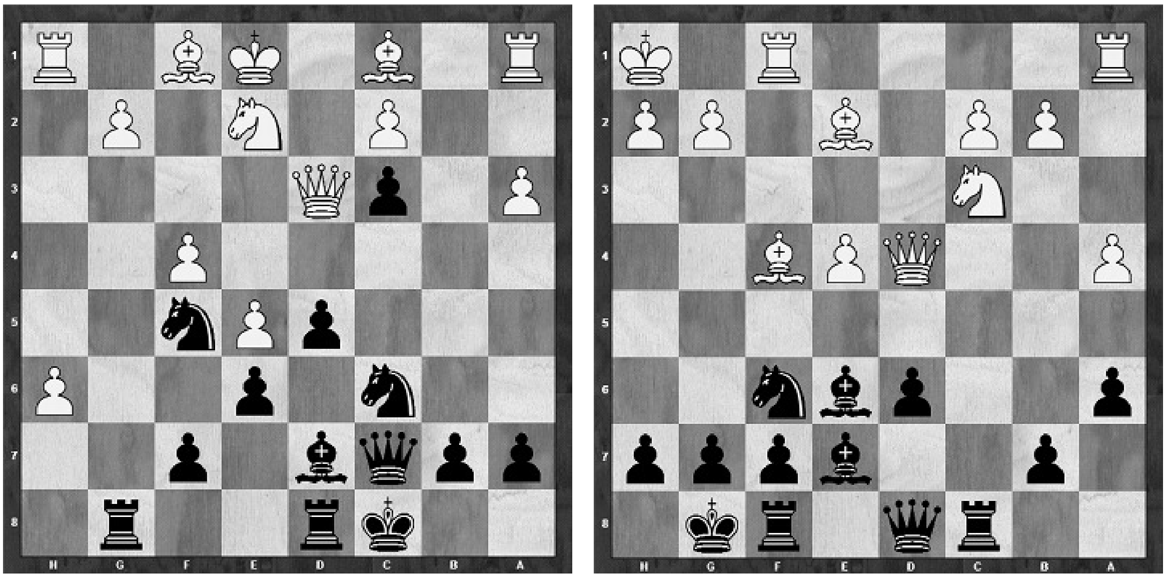


Figure 7



French position (1… Rg6)

Sicilian position (1… Ne8)

Figure 8

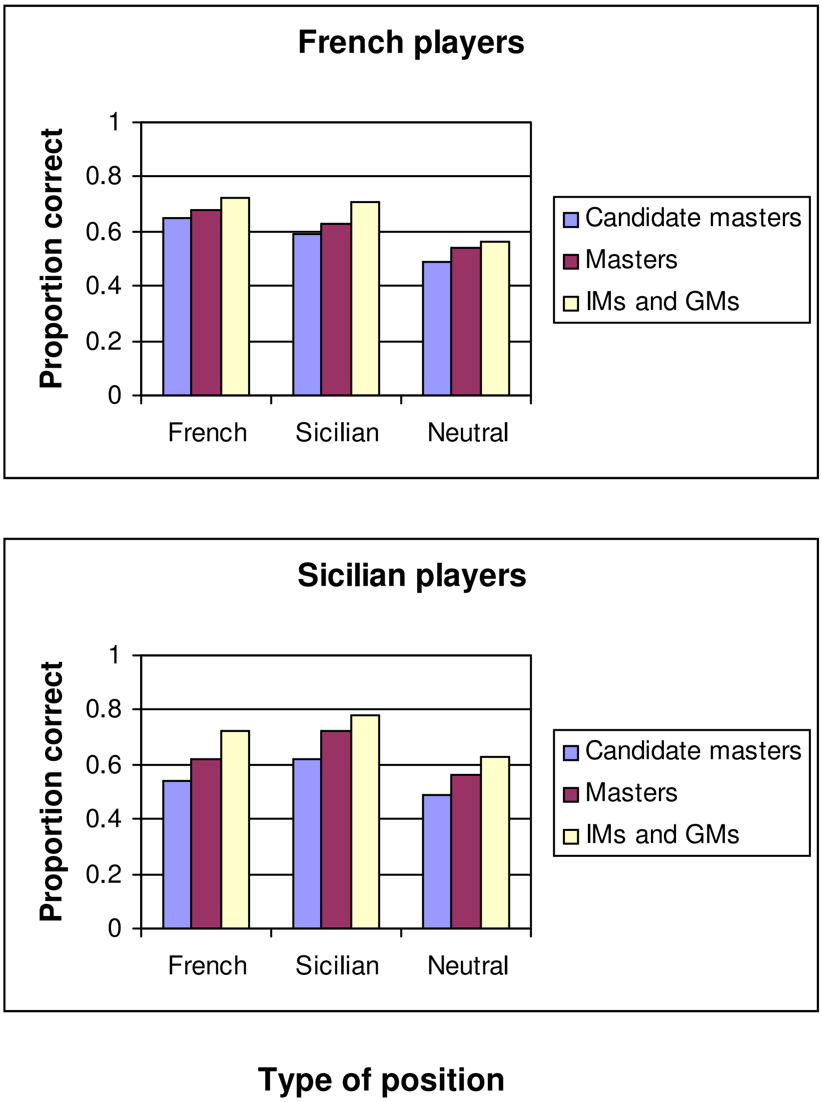
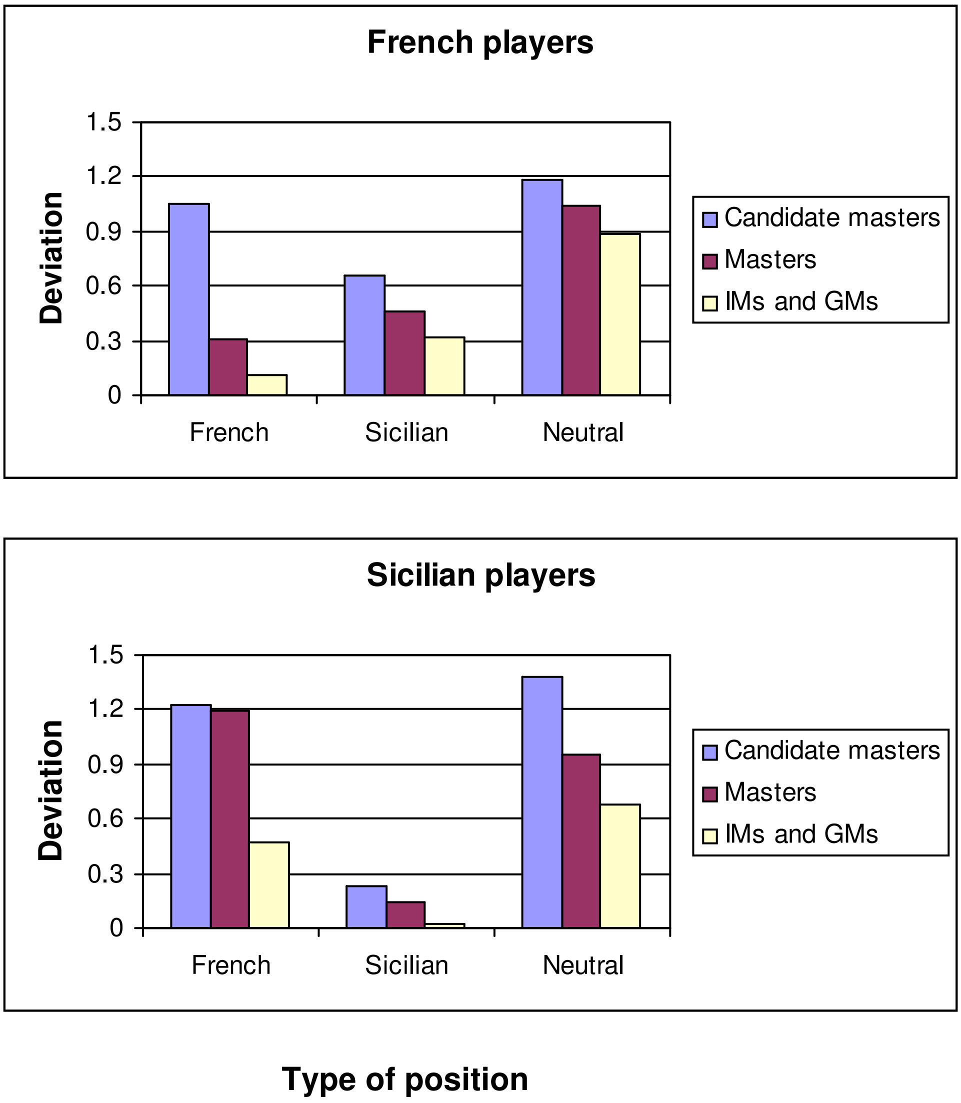

French players
Proportion correct
1
0.8
0.6
0.4
0.2
0
French
Sicilian
Neutral
Candidate masters
Masters
IMs and GMs
Sicilian players
Proportion correct
1
0.8
0.6
0.4
0.2
0
French
Sicilian
Neutral
Candidate masters
Masters
IMs and GMs
Type of position

Figure 9



French players

Sicilian players