2016

Knowledge Structures and Decision Making in Chess

Jerad Moxley

FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

KNOWLEDGE STRUCTURES AND DECISION MAKING IN CHESS

By

JERAD H. MOXLEY

A Dissertation submitted to the
Department of Psychology
in partial fulfillment of the
requirement for the degree of
Doctor of Philosophy

2016

Jerad Moxley defended this dissertation on March 22, 2016.
The members of the supervisory committee were:

K. Anders Ericsson
Professor Directing Dissertation

Martin Fennema
University Representative

Neil Charness
Committee Member

Richard K. Wagner
Committee Member

Frank Johnson
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements

ACKNOWLEDGMENTS

I am very appreciative of each person who agreed to be on my committee. Their time was very valuable and I appreciate all the time they spent on this project. I also appreciate every person who involved me in their research in anyway. Those include Sarah Wood, Natalie Sachs-Ericsson, Anders Ericsson, and Dustin Souders. I appreciate the advice of Ellen Berler, Orienda Johnson, Mark Licht, Anders Ericsson, and Tyler Towne on various teaching matters. It was an honor to be able to work with and consult with Neil Charness. His diligence, integrity, kindness, and patience are something I aspire to. No matter what happens in my life I will always remember the wisdom of Anders Ericsson. Being able to talk to him was the greatest educational experience I can imagine.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Research on the cognitive basis of chess skill has focused primarily on chess knowledge structures and their relevance in accounting for superior recall of briefly presented chess positions by chess experts. The direct relevance of research on short-term recall of chess position to superior move selection, is based on two theoretical assumptions that have not been fully tested. The first assumption is that the chess knowledge structures are automatically activated during the perception of a position in a similar manner for both tasks requiring short-term recall and selection of the best move. Because experts are thought to have greater knowledge and this greater knowledge is thought to be activated automatically it has been argued that experts are less harmed by reduction in available time for selecting a move than weaker players. The second assumption is that the knowledge structures that mediate the selection of a chess move can be adequately captured by a memory task for briefly presented positions. This dissertation attempted to test these two assumptions. Study 1 found no general interaction between time given to solve a chess problem and chess skill on the strength of move selection, instead finding a relatively stable expert advantage across times. Additionally, it established this even for the 5 s condition typically used to study memory. Study 2 presented chess positions to players for 5 s either to select the best move or to recall the position. In the condition involving the selection of a move, players were asked to recall the position after announcing their selected move. The study found substantial differences in the structure of recall between the two conditions. Additionally, certain expected features such as a bimodal distribution of response times were not found calling into question some assumptions of chunking theory.

INTRODUCTION

The most influential contemporary decision-making theories argue that expert intuition develops from years of experience with domain-specific tasks. This experience is encoded in memory in such a way that it allows a person to gain rapid access to appropriate responses to encountered situations in the domain of expertise (Kahneman & Klein, 2009). Research on chess has shown that chess players develop knowledge representations that allow them to rapidly perceive better chess moves than less skilled players (de Groot, 1946/1978). The two most important theories of chess knowledge are Chunking Theory (Chase & Simon, 1973) and Template theory (Gobet & Simon, 1996b). Both theories assume that chess knowledge is based on the storage of patterns of chess pieces located on specific squares on the chessboard. Template theory generalized the encoding mechanisms of chunking theory by allowing for chunks to develop into schemas containing slots which can rapidly encode variable pieces and variable squares (Gobet & Simon, 1996b). Both template theory and chunking theory share the assumption that the recognition of patterns of chess pieces occurs automatically upon presentation of a chess position. If this is the case, then any chess stimuli including a chess position presented for immediate recall should access the same knowledge structures as during a task requiring move selection. Also, the knowledge structures accessed with briefly presented boards are assumed to be the same as those accessed with longer presentation times. Not all theories share these assumptions which largely went untested until now. For instance, long-term working memory theory (Ericsson & Kintsch, 1995) argues that experts adapt their memory for the specific requirements of a task.

This dissertation has three main goals. The first goal is to begin to generate a shared theoretical framework for evidence about both chess knowledge and chess move selection. The proposed studies in this dissertation will extend the existing body of evidence about the effects of time to select moves for a chess position as a function of different levels of chess skill and the difficulty of the selection task. The second goal is to test the long-held assumptions discussed in the previous paragraph that expert knowledge relevant to a chess position, is activated identically independent of the task conditions. The final goal of the proposed studies will be to update existing theories of expertise based on the development of a common theoretical framework for knowledge and search.

Chunking Theory

De Groot (1946/1978) argued that perceptual skill played a critical role in chess skill after more skilled players showed extremely large memory advantages for briefly presented chess positions. A chunking theory of expertise (Chase & Simon, 1973; Simon & Chase, 1973) specified units of information created by grouping together other pieces of information, called chunks, as the knowledge structures that defined experts' skilled pattern recognition. Chase and Simon (1973) showed that a master player could recall more pieces in briefly presented game positions than a strong player or a non-player. However, this advantage on real positions did not generalize to random positions. Chase and Simon (1973) argued that the reason for this performance dissociation was that that players would not have stored chunks for random piece configurations. Simon and Chase (1973) hypothesized that the chunks had a pointer to information in long term memory (LTM) that contained information such as successful moves associated with that chunk. Because in chunking theory skill came from more and bigger chunks, the size of chunks identified using recall or the copy task are taken as the main factor that distinguishes skill. Chunking theory predicts equal capacity (as measured by number of chunks) but different sizes of chunks.

Analysis of response latencies during chess recall also seemed to establish that the master's recall advantage was due to his grouping more pieces together into larger chunks. Chase and Simon (1973) videotaped a chess player copying a chess position and found that he generally took more than two seconds to place a piece when he glanced back at the original board. They additionally found that two successively recalled pieces after a glance at the original board appeared to be less related to each other in terms of piece color, proximity, type of piece, and if the two pieces were attacking or defending each other than two successively placed pieces recalled with no glancing back the board in between their placement on the board. Chase and Simon (1973) therefore identified the memory process of recalling a piece within a chunk as occurring in less than 2s while the processes involved in remembering a new chunk took longer than 2s. The psychological reality of chunks on the recall task identified using the 2s criterion was also supported by there being more relationships between two consecutively recalled pieces within a chunk as opposed to two consecutively recalled pieces in different chunks. Gobet and Simon (1998a) further supported the identification of a process dissociation by response times by showing two very different response time distributions for within-glance and between-glance placement of pieces. Chunking theory therefore assumes that two distinct processes mediate recall and that these processes can be distinguished using response times. The first process takes longer and consists of recalling a chunk and the second process is a relatively fast access of pieces within the retrieved chunk.

Chunking theory posited how expert performance could be understood relying only on the normal capacity of short-term memory and its associated processes. According to Chase and Simon, (1973) chunks are acquired gradually in response to exposure to chess positions. Chunks are stored in long-term memory and then accessed when cued perceptually by a chess position.

Therefore, varying chess tasks, such as the copy task, the recall task, and selecting the best move could all observe the same underlying patterns and knowledge structures. After Simon and Chase's (1973) theoretical explanation of superior memory, the view that fast pattern recognition processes were the main processes that differentiated experts from novices became prevalent (Holding, 1985). This prediction has been tested many times since Chase and Simon. One of the more important questions studied since Chase and Simon has involved how experts are able to select better moves for chess positions. In particular, if expertise is primarily based on immediate recognition that would suggest relatively less reliance by experts on slow serial processes such as evaluation and search of alternative moves.

The Importance of Rapid Processes for Move Selection

Simon and Chase (1973) argued that with experience an increased number of more complex chunks were acquired which in turn mediated a superior performance. This view has been interpreted to mean that expertise is defined by rapid processes. Recognition is thought to be a process that occurs too rapidly to allow the tracking of other mediating cognitive processes if they occur (Ericsson & Simon, 1993). If rapid recognition of pre-stored chunks is the dominant factor in chess skill, then what patterns are recognized should be invariant to chess task and better players should have an increased advantage as time decreases. Although chunking theory and the related theory, template theory, do presume a contribution for search, because it allows more time for pattern recognition, both presume that immediate recognition plays a much larger role in chess skill (Gobet & Simon 1996c). One of the most important extensions of this argument was the recognition-primed decision-making model of Gary Klein (1998). According to Klein experts will recognize an event as similar to something encountered in the past and this recognition is the most important part of skilled decision making. Klein's model is similar to chunking theories, while it emphasizes fast recognition, it also incorporated search and named it *mental simulation*. In mental simulation the intuitively generated option is simulated in the mind's eye to find a flaw. If a flaw is not found the option is accepted. If a flaw is found another option is intuitively generated and mentally selected. With more experience and proper feedback the intuitively generated options should become more and more accurate (Kahneman & Klein, 2009). According to Klein (1998) increasing skill diminishes the role of search.

If Klein (1998) and Chase and Simon (1973) are correct and high levels of skill are defined by rapid recognition, then reduced time to select a given move should be less detrimental to experts than less skilled players. Calderwood, Klein, and Crandall (1988) conducted the first study testing this hypothesis. They had a small group of chess masters play a series of games against each other at either blitz speeds or normal chess speeds. They then compared their performance to a group of class B players who conducted the same task. In this study each move was rated for quality and complexity by a grandmaster. A simple move was operationalized as a move where there was one clear best move and a complex move as one where there were multiple possible moves. They found no skill differences in move strength for simple moves. For the complex moves they found that there was were no skill differences in their rated move strength in games played at normal time. However, that Masters group's moves were rated much better than the class B players moves in fast games.

Other studies of chess game performance have also been used to support the hypothesis that better players are more able to select good moves under time pressure. For instance, a study comparing the quality of a world champion's play in games with normal time limits to games with a tenth of the time found arguably a smaller decrease in performance than expected (Gobet & Simon, 1996c). Another archival study showed stronger players performed closer to their expected rating for quick games played against players rated lower than themselves than weaker players did when playing against players rated lower than themselves (Burns, 2004). These findings suggest that as skill increases the available time becomes less important in selecting high quality moves.

Archival research of the type reviewed above has the weakness that search and deliberation must be measured inductively from the results of completed chess games. However, playing a complete game has many components other than intuition about and search for moves. For instance, in the beginning of chess game better players can play better based on their superior opening knowledge (Chassy & Gobet, 2011b) and their knowledge about openings allows them to keep selecting moves deeper in the game based on direct memory recall. A more direct test using think-aloud protocols by Moxley, Ericsson, Charness, and Krampe (2012) was performed to discover if experts benefit less from additional time. . Moxley et al. (2012) adopted Kahneman and Frederick's (2005) definition of an intuitive decision as a decision that arises immediately and is adopted without change after further deliberation (see also Witteman & van Geenan, 2010 for a similar definition). Moxley et al. (2012) assumed that the move that was first mentioned in verbal protocols during the task of selecting the best move would best reflect the move generated by intuition. When players kept thinking about the selection of the move they often changed their preferences for the best move and this change was attributed to the effects of deliberation. The study demonstrated that better players selected stronger initial moves than less skilled players, a finding that is consistent with superior intuition of experts. Theoretical claim that experts rely comparatively less on deliberation than weaker players (Dreyfus & Dreyfus, 1986; Klein, 1998) and more on intuition, was not supported. Experts in this study (Moxley et al, 2012) usually changed their initial move and in general selected a clearly superior final move after search. The less skilled tournament players similarly improved their selection from initial to final move.

When examining the effects of chess expertise on the relative roles of intuition and deliberation all studies made the primary assumption that a more intuitive answer is given when the answer is produced in less time. Consequently, in Moxley et al. (2012) the first move option verbalized is considered the most intuitive. There are several problems with that assumption. For instance, in protocol studies of chess with very strong chess players, the first verbalizations often reflect broad analysis of different aspects of the chess position (such as weaknesses of the positions and attacking potentials) before verbalizing a particular move (de Groot 1946/1978). Confounds associated with free exploration of a single problem and different playing skill can best be addressed by using within-participant manipulations such as varying the time to select the best move. The hypothesis that search becomes less important as skill increases deserves further study despite the uniform theoretical predictions of Dreyfus and Dreyfus (1986) and Klein (1998), and arguably chunking theory. The empirical evidence accounted for by chunking theory is impressive, but there are important bodies of evidence that chunking theory is unable to explain. These deficiencies raise theoretical concerns as chunking theory has to date provided the most comprehensive theoretical explanation of expert intuition. These deficiencies and limitations of chunking theory eventually led to an extension of chunking theory named Template theory.

Template Theory

Chunking theory explained a large amount of chess skill data, particularly in the case of recall performance (Gobet & Simon, 1998a). However, researchers produced data that chunking theory could not explain. A key claim of chunking theory was that the expert advantage in memory was based on holding larger patterns in short-term memory (STM). Asking players to perform a task that requires the use of the limited STM immediately after studying a board should displace any chunks associated with the chess recall task from STM. Testing players of differing skills with or without an interpolated task tested the assumption of exclusive storage of the chess position in STM. Two studies found very small decrements in recall among expert chess players after completing an interpolated task (Charness, 1976; Frey & Adesman, 1976). This was true even when the interfering task involved the encoding and immediate recall of another chessboard (Charness, 1976). Small effects of interpolated tasks suggested rapid storage of the presented chess positions in long-term memory for experts.

Gobet and Simon (1996b) developed template theory to keep the core strengths of chunking theory. Templates can be analogized to a schema as they contain general plans and goals gained from prior experience. Templates have all the features of Simon and Chase's (1973) chunks such as the idea of the template having pointers to moves. Templates in the model develop from a player being repeatedly exposed to a chunk. Templates add to traditional chunks slots which allow for the rapid storage of variable information. Slots are an incorporation of retrieval structures, which is a memory structure that allows the rapid storage and retrieval of information in long-term memory by use of an associated cue. These associations are typically deliberately created (Chase & Ericsson, 1982), but in template theory they are acquired incidentally through exposure to patterns (Gobet & Simon, 2000). Slots in template theory allow either different pieces to be rapidly recalled on a fixed square or a specific piece to be rapidly recalled on differing square. A slot can also contain another chunk, a plan, and an opening system (Gobet & Simon, 1996b).

Because a single template can sometimes be large enough to support most or all the recall for a presented game position, template theory can address retroactive interference effects and multiple board studies. This is because short-term memory is not as easily exceeded with as few as one template stored for each board. Additionally, the templates incorporation of retrieval structures using slots, allows more rapid stable storage of new information into long-term memory than traditionally thought possible (Cooke, Atlas, Lane, & Berger, 1993; Gobet, 1996b, Gobet & Simon, 1998a).

Template theory is proposed to offer an integrated framework to organize intuitive and deliberative processes (Chassy & Gobet, 2011a). Several key assumptions of template theory are the same as those of chunking theory. Both assume that pre-stored patterns of chess pieces are recognized automatically and solutions are suggested automatically by this recognition (Campitelli, Gobet, Head, Buckley & Parker, 2007; Gobet & Chassy, 2008). The second shared assumption is that because knowledge activation is an automatic recognition process then the knowledge activated by chess stimuli will be invariant to the goal of the task. In other words, the same chunks that allow better performance in the recall task also guide search and allow stronger move selection Additionally template theory predicts a more important role for fast processes as opposed to search by skilled players (Gobet & Simon, 1998b).

In chunking theory and template theory it is the stored patterns in long-term memory that mediate expert performance. Evidence for both theories comes from tasks that do not involve the selection of moves. Template theory was proposed to account for Chase and Ericsson's (1982) work on skilled memory by using larger patterns and retrieval structures to explain multiple board recall. Chase and Ericsson's (1982) work argued for the development of domain-specific cognitive structures that supported the manipulation of relevant information in working memory. This is in contrast to chunking theory, which hypothesized the reliance on only normal memory structures (chunks and schemas) where skill is mediated by a greater number of more complex stored chunks and schemas. Ericsson and Kintsch (1995) attempted to extend this idea into a general theory of expertise.

Long-Term Working Memory Theory

Chase and Ericsson (1982) in their skilled memory theory, stated that expert superior memory on tasks such as digit span was based on the use of retrieval structures. These retrieval structures were deliberately acquired through training to allow the organized retrieval of digit groups by encoded associations to well learned information such as running times or important dates. An example is the method of loci whereby the retrieval structure is a sequence of familiar locations, such as a path through familiar house or park. At each location information such as the encoded groups of digits are associated with the specific locations. To retrieve the information the person simply retakes the trip mentally retrieving each encoded digit group associated with each location. Long-term working memory (LTWM) theory (Ericsson & Kintsch, 1995) retains the mechanisms of skilled memory theory, but generalizes the types of mechanisms described in the proposal for skilled memory and also generalizes these mechanisms to domains where the task is not an explicit memory task, including chess playing. The main memory demand during chess playing concerns the need to maintain a representation of the position so that it can be manipulated during search among alternative moves or sequences of moves. Presumably if maintaining and/or manipulating a chess position in memory requires all available attention resources then there will be no resources available to make plans which would make search inefficient. Skilled chess players have gained the ability to maintain memory representations of the current and future chess position while generating and selecting moves, which are rated highly by human judges or chess-playing computer programs. The ability to maintain a board representation mentally was best demonstrated by Saariluoma and Kalakoski (1998) who did not find a decrease in move strength under conditions designed to simulate blindfold chess for strong chess players. A study of elite tournament play using computer analysis also demonstrates only a small increase in number of poor moves under blindfolded playing conditions (Jeremic, Ukmirovic, & Radojicic, 2010).

Ericsson and Kintsch (1995) argued that, for example, when a skilled reader reads a text they cannot be merely relying on storage in short-term memory, which they refer to as Short-Term Working Memory (STWM). If participants relied only on storage in STWM they would lose access to earlier presented pieces of information in the text when they continue reading and encounter new information in the text. Only in the cases that the information had been in attention long enough for a new chunk to be formed in long-term memory could traditional chunking theory account for the rapid accumulation of knowledge. Even in this case the new chunk might not be retrievable when relevant for comprehending of the rest of the text. Traditional theories of memory have a difficult time explaining how chunks stored in long-term memory could be available for immediate retrieval, whenever the associated information is relevant. LTWM posits that retrieval structures allow for this by allowing rapid retrieval of information in long-term memory by associating that information with a retrieval structure created as a consequence of the specific practice activities that allowed the acquisition of high level performance.

LTWM theory encompasses skilled memory principles and the mechanisms of the construction integration model of text comprehension (Kintsch, 1988). In the construction integration model knowledge of the current state described in narrative texts is represented as a situation model. During reading, the situation model is continuously updated by integrating the incoming information in the remaining text. Different situation models can be generated by different readers because of their idiosyncratic knowledge in LTM and thus lead to a different memory representation of the text in LTM. A more refined situation model leads to a more accurate and complete comprehension of the text (Kintsch, 1998). Information from the situation model maybe needed at any point, so a large role for LTWM with its organized use of retrieval cues is to anticipate these needs and retrieve the relevant information rapidly when needed (Ericsson & Kintsch, 1995).

In chess, a chess expert starts a game with a representation of chess appropriate to the openings that the chess player is planning to use. As the game progresses the player's representation of the game will be further refined by new information from the specific moves that the player and his/her opponent selects and from information generated during exploring move-countermove sequences generated by forward search. The accuracy and amount of information acquired during forward search depends upon the ability to hold, manipulate, and interpret the hypothetical chess positions encountered during forward search. These memory demands are greatly facilitated by the use of retrieval structures to rapidly and reliably encode and retrieve information in memory. Both search and actual changes in chess positions can improve the situation model of the current game and a superior situation model of the current game should allow a skilled chess player to use appropriate tactics and strategies to best maximize his/her relative advantage. A study by Schultetus and Charness (1999) tested memory for random positions before problem solving and after problem solving. They found little expert advantage when recalling a briefly presented board without having problem solved but did find a large advantage after five minutes of problem solving. This supports that the problem solving process, as well as potentially time allows experts to create new memory representations.

Long-term working memory theory does not share the assumption of the other theories reviewed in one important aspect. In LTWM the structure of memory is created based on the structure of training in the domain, and then that structure is used based on the representation of the current task (Ericsson & Kintsch, 1995; Ericsson & Moxley, 2013). Because LTWM assumes experts attempt to adapt their memory actively to the tasks' needs based on prior practice, it does not make the assumption that other reviewed models make, that the same memory structure will be activated the same way regardless of task. The prediction of LTWM is that the more similar a task is to task performance or the deliberate practice that allowed the acquisition of task performance, the more similar the representation used will be to the representation that supports the skill.

The Current Research

The current research is designed to test two underlying assumptions of immediate recognition theories of chess. The first is that rapid processes become increasingly important as chess skill increases. The associated hypothesis is exposure to chess positions becomes more brief, differences in move selection quality will increase between players differing in chess skill (Burns, 2004, Gobet & Simon, 1996c). When additional time is provided for move selection it mainly allows weaker players to use search to improve the selection of their moves (Calderwood, et al., 1988; Dreyfus & Dreyfus, 1986; Klein, 1998). While many studies are taken as evidence for this type of interaction between chess skill and time for move selection, only two studies have tested it. Those two studies were Calderwood et al. (1988), which found a large significant interaction, and Moxley et al. (2012), which found no significant interaction. Both of those studies had limitations when testing the interaction between chess skill and time for move selection. For instance, Calderwood et al. (1988) examined moves generated during the course of a complete chess game between chess players of a similar skill level. They offered no evidence that the chess positions generated in the games between more-skilled and less-skilled players were comparable. Additionally, the measurement of move quality by the chess masters seemed to be inadequate because there were no differences between masters' and class B players when they played under regular conditions. The grandmaster's ratings of move strength only found a decrement in playing strength in games played by class B players when tested under high time pressure. Finally, the rated moves were not independent and one incorrectly selected move under the time-pressured condition might have led to a series of poor moves that were all included in the computed averages.

Moxley et al (2012) did not experimentally manipulate the time available to select moves for the presented chess positions. Instead Moxley et al (2012) analyzed think-aloud protocols verbalized while selecting a chess move. The analysis compared the quality of moves verbalized in the beginning of the selection process with the final move announced by the player. It cannot be ruled out that if the players had actually been forced to select a move very rapidly they might have selected a different move than the move they first verbalized in the think-aloud protocols knowing that they had plenty of time. The current Study 1 provides a much more rigorous test of the first hypothesis by using the well validated best-move task and experimentally varied the time limits from the 5 s traditionally used on the recall task to 180 s, which is the average time per move in tournament chess games (Gobet & Charness, 2006). Additionally, computers were used to score the generated moves for the chess positions, which should give an unbiased, reliable, and sensitive measure of all generated moves.

Testing chess players on the best move task at 5 s will also allow this paper to establish the parameters necessary to test the second assumption: that chess knowledge is rapidly activated regardless of the task assigned to the participants. To test this assumption, it is necessary to know how quickly chess players can generate accurate selections of chess moves. Once the necessary presentation time for accurate selection of moves has been determined in Study 1 the second study (Study 2) will assess the representation of the chess position that mediate the move selection.

STUDY 1

Studies have shown that stronger chess players can select the best move for an unfamiliar chess position more effectively than weaker players under time pressure (see Campitelli & Gobet, 2004, for a mention of five second time limit, and van der Maas & Wagenmakers, 2005 for a 30 s presentation time). No study has, however, compared the quality of move selection by players differing in chess skill when the time permitted for move selection is experimentally varied. Study 1 tested if better chess players gain a reliable advantage in move selection in a few seconds, as has been claimed (Klein, 1998). It will be of particular interest to identify the shortest time interval where a monotonic relationship of chess skill and move quality is found across the measured range of skill. This study will test three hypotheses (one subsumed by an interaction).

1. As available time to select the move increases, it is hypothesized that the quality of selected moves would improve.

2. Based on research studies by Calderwood et al. (1988) and Burns (2004) this effect should be qualified by an interaction whereby more skilled players benefit less from increased time.

3. Based on Moxley et al. (2012) this effect should be qualified by a triple interaction between time for move selection, problem difficulty, and chess skill. According to this finding experts benefited relatively more from more search in hard problems compared to non-experts but relatively less in easier problems.

Studies of experts and novices typically have very large effects of expertise. For instance, in a recent meta-analysis on effect sizes in chess, Charness and Moxley (2013) estimated the relationship between chess skill rating and performance on the best-move task to be r=.76. To have .8 power to find a significant effect would only require 9 participants. However, the studies in that meta-analysis did not manipulate the available time for making a move selection within the studies. The study by Calderwood et al. (1988) may have used the most relevant manipulation as it compared ratings of moves generated with different time allotments. They estimated that the effect of speed on chess play was $\eta^2_{partial}$=.26. In order to have .8 power to replicate that effect a study would need 14 participants. Additionally, this study estimated the interaction of skill and time to be $\eta^2_{partial}$=.35. To replicate this interaction 10 participants are needed to attain .8 power. Given that Study 1 is not a replication of Calderwood et al.'s (1988) study it would be prudent to increase the sample size beyond these estimates. In fact, not only does the study use specific problems as opposed to rating moves in a game but the Calderwood effect comes from positions with two or more possible moves while this study will look at positions with a single clear best move. For the other two hypotheses there are no similar studies conducted so power cannot be estimated.

Methods

Participants

Chess players were selected from chess clubs throughout the region according to the following criteria: players must be between 18 and 45 years of age (43 years old was identified as the age when chess skill begins to decline by Roring and Charness, 2008) and must have a chess rating of at least 1000 from the United States Chess Federation (USCF) or a comparable Fédération Internationale Des Echecs (FIDE) rating. Players with ratings of under 1800 were paid 15 dollars per hour and players with ratings higher than 1800 were paid 20 dollars or at the rates they normally charge for tutoring other players. The main analysis of the current study includes the data from 25 players.

The average chess rating of the sample was 1874.40 (SD=370.41). Twenty-two of the 25 participants gave descriptive information about their chess careers. Of those 22 the average age was 29.77 years (SD=6.83), the average starting age was 9.57 years (SD=6.15), the average age of beginning serious chess play was 16.98 years (SD=5.46). This sample had engaged in a very limited amount of paid tutoring, averaging only .31 years (SD=.68). In this sample years of serious chess experience was highly correlated with chess rating $r(21)=.77$, $p<.001$.

Design

Participants were instructed to select the best move for chess positions under varying time constraints. Before every trial the chess players were told how much time would be available before they had to announce their selected move. There were five different durations for completing the selection of a move, namely 5, 10, 20, 30, and 180 s. These presentation times were selected for the following reasons. The traditional time for recall of chess positions is 5 s. The shortest presentation time in previous research on move selection that demonstrated a reliable effect was 30 s. Analyses of move selection (Gobet & Charness, 2006) in tournament games suggest an average time for selecting moves was around 180 s. The other intermediate presentation times were selected to allow assessment of the shortest presentation time that would show a reliable relation between chess skill and move selection. There were 10 trials for each presentation time for each participant. During each trial of move selections the players were asked to think aloud. In half of the trials for the 5 and 10 s condition the participants were asked to give a retrospective report of their thought processes (Ericsson & Simon, 1993) after they reported their selected move. A trial was concluded by asking participants if they remembered having seen this presented position prior to the start of the current experiment.

Materials

Positions were presented on a laptop with a 13-inch screen using the PowerPoint program. Participants were recorded thinking aloud using video capture software. Move selections were recorded orally. Positions were selected from real chess games according to the following procedure. The position had to be a critical position with a clear best move. A critical position with a clear best move was defined as a position where the bests move as measured by Rybka 3 was at least half a pawn better than the second best move. This position had to have occurred in the game immediately after the opponent had made a mistake of at least a half pawn value as judged by the same chess program. This particular operationalization was motivated by a computer analysis of one of the famous problems studied by de Groot (1946/1978; Problem A). This problem met the criteria applied in this study.

Procedure

The study began with a short instruction with warm-up examples of the procedure for giving verbal protocols (Ericsson & Kirk, 2000). Then they were given a set of 4 practice problems. Two of these problems were presented at 5 seconds and the other two problems were presented at 10 seconds. Before the start of every trial the computer screen displayed the amount of time they had available to generate their move selection and this was followed by the display of the problem. The participants announced their selected move verbally. During the practice problems the experimenter emphasized the importance of announcing the selected move at or before the end of the presentation time. The computer screen was programmed to alert participants to how much time remained before a move had to be announced. When less than 4 seconds remained the peripheral areas of the computer screen turned yellow and when less than two seconds remained these areas turned red.   Following the practice trials the experimental session began. The 50 problems were randomly assigned to one of 5 problem sets consisting of 10 chess positions. The five sets were rotated across the 5 experimental conditions with different presentation times using a Latin square design. When the problem sets were assigned to presentation conditions, the order of problems were determined randomly.

Description of Measures of Accuracy of Move Selection

Computer analysis was performed on each problem by Rybka 3 to score the quality of the participants' selected moves. Chess positions can be extremely complex and are not designed to be aggregated into a psychometrically valid composite. One example of a difficulty in analyzing chess problems is that it is impossible to know beforehand what moves will actually be selected. This means that the two problems that may appear superficially similar in the range of computer scores of possible moves may not behave that way in practice. For instance, if both positions have an extremely disastrous move but players never select that moves in one case but do in another then one problem may have a much higher variance than the other. This can become an issue when an attempt is made to aggregate multiple problems. For instance, if we score checkmate as being a 40 pawn advantage (more of an advantage than all other pieces combined would be) there could be a problem where some participants select the best possible move and this move leads to winning by mate for a score of 40 points. If for the same problem many other participants select the worst possible move which leads to white losing by mate this problem would have a range of 80 pawns. Another valid problem could have a range as little as .5 pawns if all moves chosen but the best one were equal. If these two problems are aggregated almost all of the variance would be accounted for by the first one, which would eliminate much of the advantage of aggregating many problems. Most robust estimating procedures would have almost the inverse problem, for instance a trimmed mean would mainly eliminate responses from the first problem meaning the second would dominate. Therefore, I will use a form of Winsorizing which will be based on chess pawn points as opposed to the statistical standard deviation. This will be done such that the best move is exactly .5 pawns better than the second best move for each position and no move is more than .5 pawns worse than the mid-ranked move and/or 2 pawns worse than the best move. This means each problem will have a range of a minimum of .5 pawns separating the best move from the worst move and a maximum of at most 2 pawns difference between the best and the worst move.

Analysis

The main analysis was a mixed model regression with random effects assigned to the individual subject for intercept and the slope of time of problem within person. Additionally, since tasks vary by the presentation time a model was tested allowing for random effects of problem and a random effect of the slope of time within problem. A model comparison approach was taken. Models assessed the need for both random effects of persons and problems. Three models were compared, namely a linear model of time, a log linear model of time, and a factorial model of time. The best fitting model was interpreted. Mixed models provide very flexible forms of analysis and have a couple of advantages for this study. First, it was of theoretical interest if there was heterogeneity in how chess players respond to both problems and to time pressure. Random effects allow clear testing of these hypotheses. Additionally, mixed models can make more realistic assumptions about the error covariance structure of the condition cells. In this study the cells may have different reliability and hence the correlations of conditions may not be equal. As to fixed effects the skill factor was treated as a continuous variable and allowed to interact with time. This analysis tested if move strength increases with time and if the skill by time interaction was present such that better players have the largest performance advantage with the shortest periods for move selection. This model will be tested for both Winsorized score of move quality and if the selected move was the best move or not. This model will be elaborated by adding a factor measuring difficulty for selecting the best move for a given chess position. This variable will be computed as the averaged predicted value for each move as computed by a random effects model identical to the analysis above but instead analyzing if the move was correct or incorrect. This will allow difficulty to be measured for each item after controlling for the skill of the player, the available time the player had to select the move and the interaction of the presentation time and the skill of the player.

Next Cronbach's alpha and interclass correlation coefficients were computed. This was done using the difference for each player on each move of the Winsorized score and the predicted value of the problem computed above.

Results

Analysis of Proportion of Announced Moves Matching the Best Move

The first analysis will be whether a player selected the best move for a given move. The linear fixed effect of rating was significant and positive $t(143)=3.46$, $p=.001$, as was the effect of presentation time $t(51)=3.52$, $p<.001$. The interaction of time and rating, however, was not significant $t(1204)=-.61$, $p=.55$. For the random effects there was a significant effect for the intercept with item $z=2.75$, $p=.006$. The random effects of players (beyond the effect of chess rating) $z=1.92$, $p=.06$, were nearly significant. The effect of the interactions between presentation time and chess position was $z=1.33$, $p=.18$. All other random effects were very small and non-significant. A predicted value was output from this analysis for each player for each position these were then aggregated for each position. This measure of difficulty should be purer than the raw score as it accounts for both the player strength and the amount of time that player had. That said, this measure correlates almost perfectly with number correct $r(23)=.997$. While this analysis was not intended primarily as a design check this basically identical correlation largely suggests that item, rating, and time were not confounded.

Analysis of Winsorized Move Scores

To analyze move strength as a continuous variable, the computer score was used. This score tended to have bad statistical properties when analyzing all 1208 selected moves (skewness=-6.31, Kurtosis=66.12). Therefore moves were Winsorized as described in the procedure section by setting the best move to be exactly .5 pawns better than the second best move (the least difference that was accepted). The worst move was likewise set to be .5 pawns worse than the mid-ranked move with a maximum range set at 2 pawns. This greatly improved the properties of the distribution (skewness=-.75, Kurtosis=1.23) and improved the correlation with rating of the aggregate measure from $r(23)=.53$, $p=.01$, to $r(23)=.74$, $p<.001$ (for correct or incorrect the correlation is $r(23)=.70$, $p<.001$). Reliability for the scale using the Winsorized data was .80 compared to .74 for the non Winsorized data. See Figure 1 for condition means by rating.

The next analysis replicates the exact same model as was presented with the correct/incorrect problems but assumes a Gaussian error term. This means the fixed effects were log-transformed time, rating, and the interaction of log-transformed presentation time and chess rating. Random intercepts were specified for the subject and the problem. Presentation time was allowed to have a random slope for both participant and problem while rating and the interaction of rating and time were allowed to have random slope for problem. Quadratic, linear, and factorial models of time were also tested but in each case the log-transformed model provided the best fit. As for the random effects the intercept of player was not significantly random, $z=1.96$, $p=.05$, but as would be expected in this case, item was $z=4.92$, $p<.001$. Again log time was not significantly random $z=1.6$, $p=.11$, and again the other random effects all rounded to zero.

Similar to the previous analysis of accuracy of selecting the best move, chess rating was positively related to move strength $t(146)=4.21$, $p<.001$, as was log presentation time $t(46)=4.58$, $p<.001$. The interaction was again not significant $t(1317)=-1.05$, $p=.30$. For the random effects the intercept of player was not significantly random $z=1.96$, $p=.05$, but as would be expected in this case item was $z=4.92$, $p<.001$. Again log time was not significantly random $z=1.6$, $p=.11$ and again the other random effects all rounded to zero. The linear and factorial models were also tested and in this and every other case provided substantially worse model fit by traditional rules of thumb (Raftery, 1995)

Next a model adding item difficulty as measured by predicted move in the analysis of correct or incorrect move was added. For this analysis based on the previous analysis I simplified the model to have only random intercepts of subject and item with both only having a random effect of log time. While none rounded to zero, again only item intercept was significantly random $z=4.92$, $p<.001$. For fixed effect, the effect of rating was significant $t(109)=4.65$, $p<.001$, as was time $t(45)=4.64$, $p<.001$. The interaction of time and rating was not significant $t(1132.17)=-1.41$, $p=.15$, nor was item difficulty $t(50)=1.47$, $p=.15$. For this model BIC=2720.92 and AIC was 2669.95.

Next all interactions were added and the model with all two way interactions was clearly inferior, with a BIC of 2732.30 and a AIC=2671.14 with no two-way interaction being significant with the interaction of rating and item difficulty being t(1118.50)=1.40, p=.16 and the interaction of log-transformed time and item difficulty being t(46.75)=1.35, p=.19. When adding the three-way interaction, the model fit statistics when comparing to the first model were mixed as BIC got worse BIC=2730.52 while AIC improved AIC=2664.26. Again the only random effect that was significant was item intercept z=4.92, p<.001. The main result for this analysis is as BIC got worse BIC=2730.52 while AIC improved AIC=2664.26. Again the only random effect that was significant was item intercept z=4.92, p<.001. The main result for this analysis is

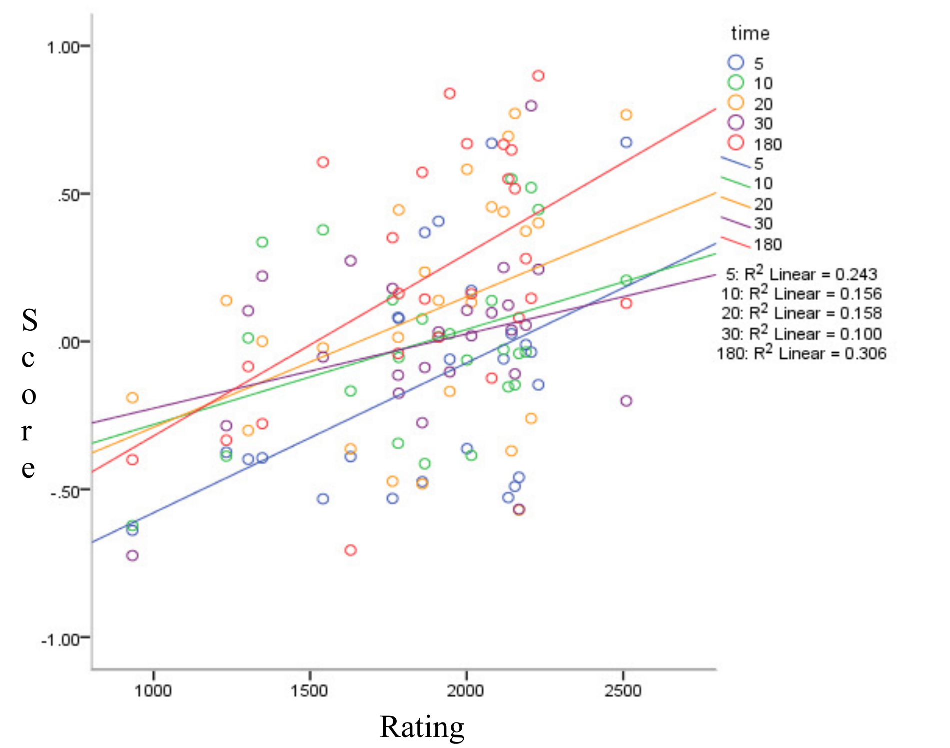

Figure 1. Relationship between rating and Winsorized move score for each condition.

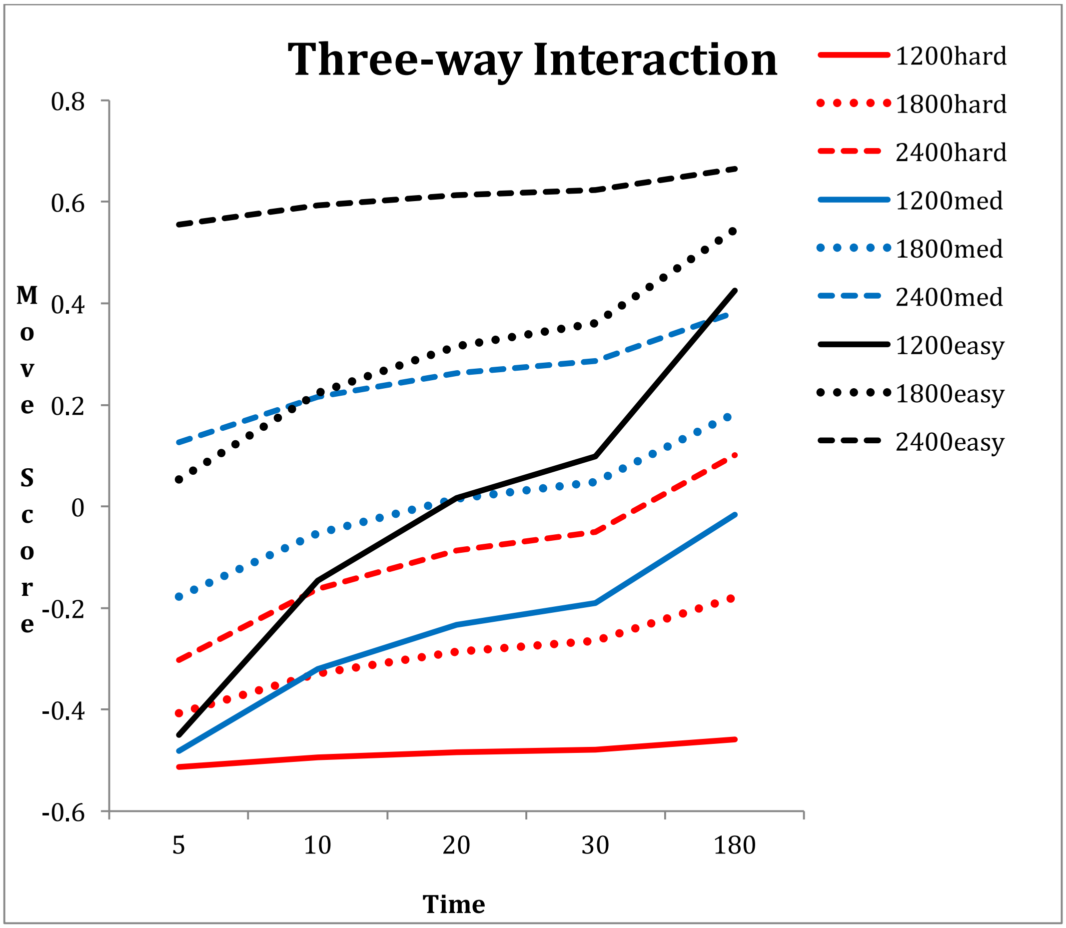

Figure 2. Three-way interaction between rating time and problem difficulty. The layout of the chart is that the easier problems are colored black, mean performing problems are in blue, and the more difficult problems are in red. Undashed lines represent the weaker players in the sample, the line with small dashes represents the approximately mean rated players and the lines with longer dashes represent the best players.

as BIC got worse BIC=2730.52 while AIC improved AIC=2664.26. Again the only random effect that was significant was item intercept z=4.92, p<.001. The main result for this analysis is that the 3-way interaction itself was significant $t(1143.28)=-2.89$, p=.004. Given that the three-way interaction is strongly significant, the reason for the contradictory model fit results must be the cost of adding the extra two-way interactions, which are not contributing to model fit. Given that the model fit is contradictory it seems reasonable to base interpretation on the significance of the parameter itself which is strongly significant.

The meaning of the three-way interaction is apparent when plotting it (Figure 2). For the easiest problems the best players benefit very little from search on easier problems while the weaker players benefit greatly. In contrast, on relatively difficult problems the weaker players benefited very little from increased time while the stronger players benefited a fair amount. Simply put on easier problems more time reduced the differences between players while on more difficult problems more time increased the differences between players.

Discussion Study 1

The results of the study found a benefit of longer times for making move selections. The experimental manipulation of time available for selection produced results consistent with the findings of Moxley et al. (2012). The hypothesized interaction between presentation time and chess skill was not supported. Both Moxley et al. (2012) collecting verbal reports and Study 1, with an experimental manipulation showed no interaction of time and skill as was strongly predicted by Calderwood et al. (1988) in analysis of moves with multiple options in full games and Burns (2004) using tournament performance. While the statistical power of this study is not adequate to establish a null result or even that an effect is small, in combination with Moxley et al. (2012) this study raises doubt about the argument that increased expert advantage with reduced decision time is a defining a feature of expertise. At the very least it appears there are specific situations such as very difficult problems, where time mainly benefits better players. There was no evidence of ceiling or floor effects. Looking at the 10 easiest problem shows that no player solved them all, the highest rated player answering 8 correctly. Additionally, for the 10 most difficult problems no player received the lowest score every time, the most being for the weakest rated player who had the lowest score for 7 out of 10 positions. No other player received the lowest score for more than 4 out of the 10 positions.

These findings are consistent with a three-way interaction between time, skill, and difficulty predicted by Moxley et al. 2012. The study found significant evidence for this triple interaction. When faced with a relatively easy problem an elite player will rapidly arrive at the solution but a weaker player will benefit from more time. However, when faced with relatively difficult problem, weaker players will not improve their selections of a move even given more time, whereas elite players will be able increase the probability of finding a better move with additional time. The proposed interaction can also provide a possible explanation for the interaction of the type predicted by Calderwood et al. (1988). It is possible that relatively simple problems, by the standard of this study and Moxley et al., (2012), would show that type of interaction and if those predominated Calderwood et al (1988) that would explain the differences. It should be noted that no problems analyzed in obtaining the Calderwood et al. (1988) interaction would have been used in the current study or Moxley et al. (2012) because their interaction was found for positions with multiple possible good moves. The problems I defined as critical positions would have been classified as simple problems by that study and for this type of "simple" problem Calderwood et al. (1988) found no effects, even failing to differentiate skill. However, Klein and Peio (1989) noted that simple problems were mainly forced moves or positions with pieces that could be captured with an immediate gain in material. Therefore the critical positions in the current study, which were never forced, would have been lumped in with many forced moves or positions with a move that would directly gain material. The fact that critical positions in this study elicited reliable skill differences, while simple positions in Calderwood et al. (1988) did not, is consistent with this argument. However, it is impossible to know which of these types of positions define skill in real chess games. Therefore, the null interaction of time and skill for all problems combined in this study may be present in chess as played. The interaction could exist either because most moves are easy or because other types of positions are more important. Additionally, the idea that in what Calderwood et al. (1988) defined as complex problems, experts found one of multiple acceptable moves quicker must also be considered a possibility yet unconfirmed.

The interaction between difficulty, decision time and chess rating suggests that the expert advantage over less skilled players is greatest for hard problems with sufficiently long time for move selection or easy problems with brief times. The current study presented isolated positions for move selection, whereas Burns (2004) and Calderwood et al. (2004) analyzed sequences of moves in entire chess games. During playing complete chess games one could speculate that a strong enough player could also control the chess game at either time frame by attempting to create simpler or more complex situations based on the amount of time in the game.

These results make Study 1 very relevant to the study of recognition and search in chess. The study adds significantly to a now growing body of research that questions how fundamental the interaction of skill and time is in chess. The current study suggests instead that there are conditions under which experts benefit from search in contradiction to predictions of pattern recognition theories. Another goal of Study 1 was to identify the shortest decision time which produced relatively consistent effects of skill for the best-move task. The significant correlation between skill and move selection even in the 5 s condition supports the idea that the knowledge representations that support chess skill are available within the 5 s usually used in the recall task. This suggests that it would be possible to design an experiment where players would either select the best move or recall the same chess positions with the identical presentation time of 5 s.

STUDY 2

A key hypothesis of chunking theory, template theory, and even experience recognition theory is that chunks can be identified by analyzing the recall of briefly presented chess positions. Almost all research on chess move selection has given players 30 s or more to select the best move. Until the findings from Study 1 it was not clearly established that significant skill effects would be present in the best move task administered with 5 s decision times. This finding permits the study of move selection with short decision times that are shorter than the times assumed by chunking theory, to be necessary for storage of new chunks in LTM. With decision times longer than 7-9 s, chunking theory would not rule out the creation of new chunks in LTM (Gobet & Simon, 1996b). The critical presentation time of 5 s has been traditionally used because 8s is not an exact estimate of the time needed to form a chunk (Simon, 1974) and thus a time limit slightly below 8s would provide better protection against storage in LTM. If skill differences in move selection were not represented by move selection within 5 s then the recall task with 5 s presentation time might not rely on the existing chunks that mediated move selection. Having established that the creation of new chunks is not a problem it is possible to test the hypothesis that the same chunks drawn from STM are involved in both the move selection task and the recall task. Chunking, template, and experience recognition theories assume the encoding of the chess position occurs automatically for chess tasks. Therefore when players are studying a chess position with the intention of selecting the best move for a chess position, after the position is removed from view they should recall the board with the same chunking features they would recall if they were asked to focus on recall only.

In order to compare recall performance when participants select the best move to when they simply recall the position it is necessary that the types of chess positions that are presented be similar. The chess positions typically used for the memory task have been quiet middle game positions, which are defined as positions about 20 moves into the game which are not in the midst of an exchange (Chase & Simon, 1973). In contrast, past studies of best move selection have used either critical positions (e.g. de Groot, 1946/1978) or positions where the best move gave a positional advantage or led to a winning position as judged by a strong chess player (e.g. Charness, 1981). Study 1 offered a more precise definition of a critical position than past studies: a position with a half pawn difference as measured by a strong program between the best move and the second best move coming after the opponent had made a mistake of at least a half pawn value as judged by a strong chess program. None of the chess positions used as exemplar stimuli in recent studies of chess recall would have been selected for Study 1, as well as likely the other studies, due to there not being enough difference between the best move and the second best move (positions tested drawn from Gobet & Simon, 1996a; Gobet & Waters, 2003, Waters, Gobet, & Leyden, 2002). Both of these types of chess positions differ from the ones selected by Linhares, Freitas, Mendes, & Silva . (2012), where the positions were defined as the final position of a game. The positions studied by Linhares et al. (2012) would not have been included in best-move tasks or recall tasks as they were end game position with a decisive advantage for one player, often with critical pieces en prise or the king in check. The fundamental differences even in the positions selected makes it very difficult to compare across studies using different designs. Study 2 is the first study presenting the same positions and the task being randomly determined. Additionally the positions in Study 1 and Study 2 were randomly selected from a common pool of possible positions.

Chunking theories as stated in the introduction predict automatic recognition of chess positions. If this is correct the goal of the task should not change the essential features of recall as measured by speed of recall, the number of pieces recalled, the size of chunks, or the number of chunks. Additionally, chunking theories identify chunks by response times. This would mean that the distributions of response times in recalling chess pieces will be a mixture of recall of chunks and pieces within chunks and thus display a bimodal distribution reflecting two separate processes. Additionally, pieces within chunks should be more related to each other than they are to pieces in other chunks. The study has the following hypothesis.

1. The null hypothesis was that the recalled pieces would not differ when the task was to recall a chess position from those of chunks that mediate the selection of the best next move. Consequently, no differences were predicted in observable characteristics of chunks when individual chess players were instructed to perform the recall task compared to when they were instructed to perform the best move task as their primary task. The studied characteristics of recall were number of pieces correctly recalled, size of the largest chunk, and number of chunks.

2. It is only possible to analyze the relation between quality of selected moves and the subsequent recall for trials when the move selection task was the primary task. If the within-subject effects for recall characteristics and for quality of selected moves were significant that would suggest the possibility that greater understanding of the position led to both better move selection and more extensive recall. This would lead to a hypothesis that as performance on move selection increased on a given board number of pieces correct would increase and in particular the size of the largest chunk would increase. The between-subject effect on recall and quality of selected moves

would point to a more general phenomenon whereby people who are doing well on one task are doing relatively better on another task.

3. Similarly, no differences in recall times for individual pieces are predicted, especially after the completion of the first and second piece. Additionally, both conditions should have equal variance and the same shape.

While difficulty of the move selection for a chess position was an important factor in Study 1 it will not be considered in Study 2. When I conducted an analysis of two-way interaction (difficulty and rating) for the 5 s condition there was no interaction present in Study 1.

Methods

Participants

Because age is likely to be a major factor influencing performance, only adults between the ages of 18 and 45 were invited to participate. Players were recruited from a wide range of chess skill as in Study 1. From the 25 players that were recruited, three participants had incomplete data and could not be analyzed in any analysis that required knowledge of rating or age. Payment structure was identical to Study 1, ratings were based on USCF or FIDE systems. The mean chess rating of the players in this sample was 1662.04 (SD=418.55) and the average age was 29.13 (SD=9.41).

Design

The study had two task conditions. On half of the trials the players were instructed to select the best move as the primary task, followed by recall of the presented position. On the other half, participants were simply instructed to study the presented position to be able to recall the positions. The primary task for a given trial was announced before the start of the trial. Based on the findings and results of Study 1 the presentation time for the chess position was selected to be 5 s.

Materials

Chess positions were selected using the same criteria that were used in Study 1 and were grouped as 4 sets of 6 positions and then randomly assigned to conditions. The four condition orderings followed an ABBA, ABAB, BAAB, and BABA format. A represents a block of six trials of the best move task and B represents a block of six tasks with recall only. The 4 sets of problems were again rotated through the conditions using a Latin square design with problem order being random. Boards were presented on a 13-inch laptop computer. The participants recalled the chess positions by using a mouse whereby they had to click on an individual piece and then click on the square they wanted the recalled piece to be placed. They had to click on a new piece even if two consecutively recalled pieces were the same piece, such as a white pawn. In the condition with the selection of the best move, participants were asked to announce the best move verbally within a second of when the board was no longer presented on the computer. Participants were trained with practice problems to be able follow this procedure.

Procedure

A short instruction period on the task and how to recall positions was given to familiarize participants with the design. Each position was presented in the following way. The procedure started with the participant being told verbally what the primary task was (recall or move selection) and then the position was displayed for 5 s. Then a black screen was displayed for 1 second. The participants were then allowed to perform the task; either recall only or selecting the best move then recall. Participants were given 4 practice items – two where the primary task was to simply recall the position followed by two tasks where the primary task was to first select the best move and then recall the position. After each trial participants were asked to tell the experimenter if they thought that they had seen the presented chess position before the experiment. Think aloud protocols were collected on all the problems according to the same procedures in study 1. In addition, participants were asked to give retrospective reports for six trials (three move selection tasks and three recall tasks).

Results

Descriptive Analysis of Recall Performance

Averaged across both tasks the mean number of pieces recalled correctly per trial was 7.80 (SD=3.83) the average number of pieces of the largest chunk was 4.22 (SD=2.00), the average number of chunks was 6.96 (SD=3.10). The mean proportion of best moves selected was .15 (SD=.16). The means and standard deviations for the main variables are given in Table 1 for each task condition. The reliability of the complete recall test was high (Cronbach's alpha=.98).

Regression Analysis of Recall and Best Move

First a series of regression analyses will examine the between-subject's differences in the two main variables, namely the average number of pieces correctly recalled and the average strength of the move selected to confirm that broadly the data is similar to past research. To control for guessing in the recall of pieces' effects of number of incorrectly recalled pieces (errors) was statistically controlled in the analysis of chess recall. The analyses will also assess the effects of age and chess rating.

The regression analysis of the number of pieces correctly recalled included three predictor variables, namely rating, age and recall errors. The full model was highly significant $F(3,19)=9.41$, $p=.001$, $R^2=.60$. The number of incorrectly recalled pieces was positively related to the number of correctly recalled pieces $t(19)=2.75$, $p=.013$, $f^2=.71$. Rating was also positively related to recall performance $t(19)=2.98$, $p=.008$, $f^2=.79$. Age was negatively related to performance $t(19)=-2.12$, $p=.047$, $f^2=.48$.

Following the methods described in Study 1 the best move score was Winsorized. The regression analysis showed that a model including both age and rating was significant $F(2,20)=5.18$, $p=.015$, $R^2=.34$. In this model rating was significant and positively related to move strength $t(20)=2.32$ $p=.03$, $f^2=.32$. Age was not significantly related to move performance $t(20)=-2.00$ $p=.06$, $f^2=.25$.

Analysis of Recall Performance as a Function of Primary Task

The effect of the primary task on differences in the pattern of recall was examined first. The three dependent variables that were analyzed were number of pieces correctly recalled, number of chunks recalled, and the size of the largest chunk. Each of these three variables were analyzed using an identical multilevel model regression with random effects of subjects, chess position, and the random slope of chess rating. Additionally, the following fixed effects were tested: whether the variable is measured at the item level or the subject level. Manipulation (item level), number of recall errors (item level), centered rating (subjects level), and centered age (subjects level) and the interaction of rating and manipulation (a cross-level interaction) were added. In each analysis the effect of task on characteristics of recall was assessed along with any interaction with chess skill.

In the analysis of the number of pieces correctly recalled, the random effect of chess position was significant z=2.65, p<.01 as was subject z=3.22, p<.01, but the random effect of slope of chess rating within problem was near zero. The results for the fixed effects showed

Table 1

Means and Standard Deviations for various variables by condition. Best move is proportion correct.

Variable	All	Best Move	Recall Condition
Correct	7.80 (3.83)	7.71 (4.18)	7.88 (3.61)
Proportion Correct	.35 (.17)	.34 (.18)	.35 (.16)
Largest Chunk	4.22 (2.00)	3.82 (1.97)	4.63 (2.14)
Number of Chunks	6.96 (3.11)	7.11 (3.34)	6.82 (2.95)
Best Move		.15 (.16)	

showed that number of recall errors predicted fewer pieces correctly recalled *t(*550)=-4.43, p<.001 as did age *t(*22)=-2.87, p<.01. Higher rated players recalled more pieces correctly *t(*22)=2.14, p=.04. The effect of task condition was not significant with the direction of the trend being for less memory in the best move condition *t(*506)=-1.31, p=.19 and the interaction of rating and task condition was also not significant *t(*509)=1.50, p=.14.

The analysis of the number of chunks revealed only a single random effect that reached significance, namely subject z=3.12, p=.002 (see Figure 3). The analysis of fixed effects showed that more recall errors significantly predicted more chunks *t(*539)=10.94, p<.001. Players recalled more chunks in the best move condition *t(*506)=2.83, p<.01. There was no significant effect of rating *t(*25)=1.84, p=.08, or age *t(*22)=.14, p=.89. The interaction of rating and task condition was not significant *t(*527)=-.17, p=.86. The final variable analyzed in this section is the size of the largest chunk. The same pattern of significant results were observed with a random

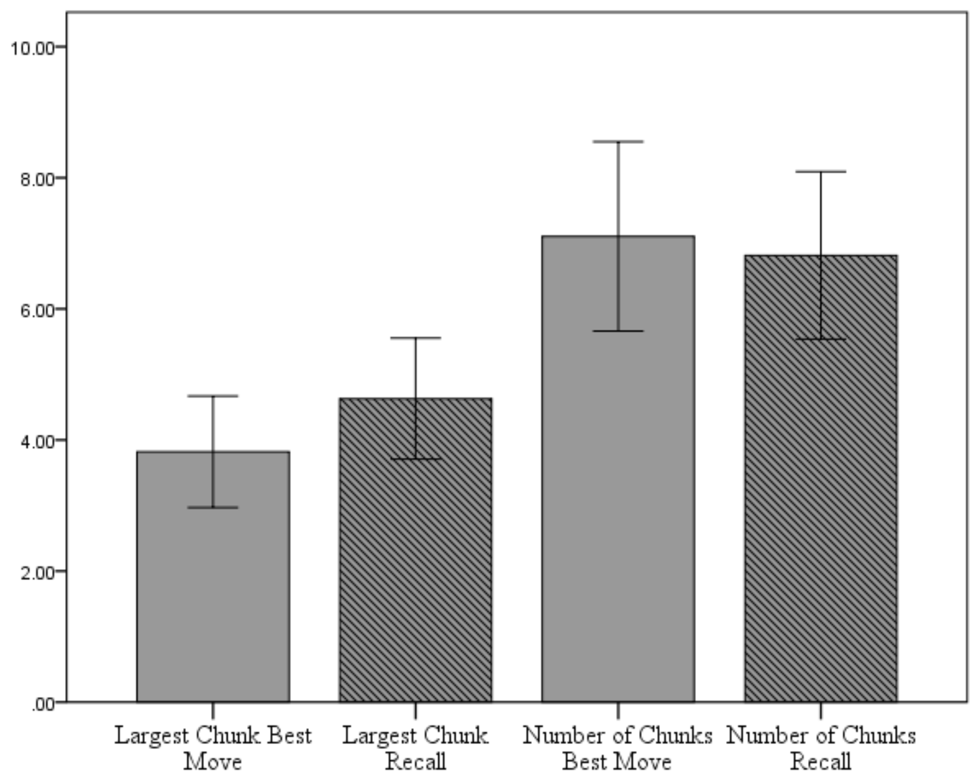

Figure 3. Means and 95% confidence interval for chunking variables in both conditions.

for subject z=3.07, p<.01 and fixed effects for more errors predicting larger chunks *t(*474)=-10.41, p<.001 as did a higher rating *t(*22)=2.28, p=.03. Older age predicted recalling smaller chunks *t(*23)=-3.81, p=.01 as did the best move- condition t*(*508)=-4.69, p<.001. The interaction of rating and task condition was not significant $t(515)=-1.75$, $p=.08$. Figure 3 provides the overall means and confidence intervals for the chunking variables by condition.

Analysis of the Relation between Quality of Selected Move and Amount of Correct Recall

Multilevel models of recall performance in best move task. The multilevel model of the best move condition had random effects of board and subject and a random slope of board by rating. Fixed effects were again errors (item level variable), rating (subjects level variable), age (subjects level variable), overall move score (subject level variable), and move score for the board (item level variable). These analyses tested the hypothesis that that the effect of quality of selected move is significant at both the between subjects and the within subject's levels. Based on the results from the previous analysis no interactions were specified. The random effect of subject ($z=2.90$) and board ($z=2.22$) were significant but the slope of rating within board was not significant. Errors was not related to recall $t(270)=-.96$, $p=.34$. Rating was also not related to recall in this analysis $t(20)=1.62$, $p=.12$, nor was age $t(21)=-2.07$, $p=.051$. The within subjects move score variable was not related to recall $t(226)=1.66$, $p=.10$ but the between subjects move score variable was positively related to recall $t(22)=2.13$, $p=.045$.

The same model was used to analyze the size of the largest chunk. The random effect of subject ($z=2.46$) was significant but board ($z=1.21$) and the slope of rating within board ($z=.88$) were not significant. Number of recall errors was related to larger chunks $t(194)=2.76$, $p=.006$. Rating was not related to the size of the largest chunk $t(21)=.52$, $p=.61$, but age was associated with a smaller largest chunk $t(20)=-3.71$, $p=.001$. The within subjects move score was not related to size of the largest chunk $t(221)=.15$, $p=.88$ nor was the between subjects move score variable $t(22)=1.80$, $p=.09$.

The same model was fit for the number of chunks. The random effect of subject (z=2.81) was significant but board (z=.22) and the slope of rating within board (z≈0) were not significant. Errors predicted more chunks $t(267)=7.48$ p=<.001. Rating was also not related to the number of chunks $t(20)=1.13$, p=.27, nor was age $t(20)=.15$, p=.88. The within subjects move score was not related to the number of chunks $t(221)=-.72$, p=.47 nor was the between subjects move score variable t(21)=98, p=.34.

Response Time Analysis

The final analysis will examine the recall times for chess pieces across the two conditions. Given the extreme skew of the distribution these times were log transformed. First, the effects of switching tasks will be examined by analyzing the retrieval times for the recall of the first two chess pieces under each task condition. These two retrieval times for the first and second piece recalled will be excluded from the remaining analysis of retrieval times. The analyses will examine if the distributions of the retrieval times differ across task conditions.

The effect of switching between move selection and recall. To study the effect of switching tasks from selecting a move to recalling pieces we analyzed the retrieval times for recalling the first two pieces. For the first piece the response times for the best move task the between-subjects mean was 3.82 (SD-.24) for the recall condition the mean was 3.49 (SD=.14). This translates to a switch cost of about 3470 ms difference and the difference is highly significant $t(21)=7.14$, $p<.001$, d=1.52. The two response times were correlated as well $r(20)=.47$, p=.03. For the second piece the mean for the best move condition was 3.34 (SD=.10) while for the recall condition it was 3.22 (SD=21). This translates to a difference of 560 milliseconds. The difference is significant again $t(21)=4.18$, p=.001, d=.89. The two response times are highly correlated again *r*(20)=.74, p<.001. For all other pieces the mean for the best move condition

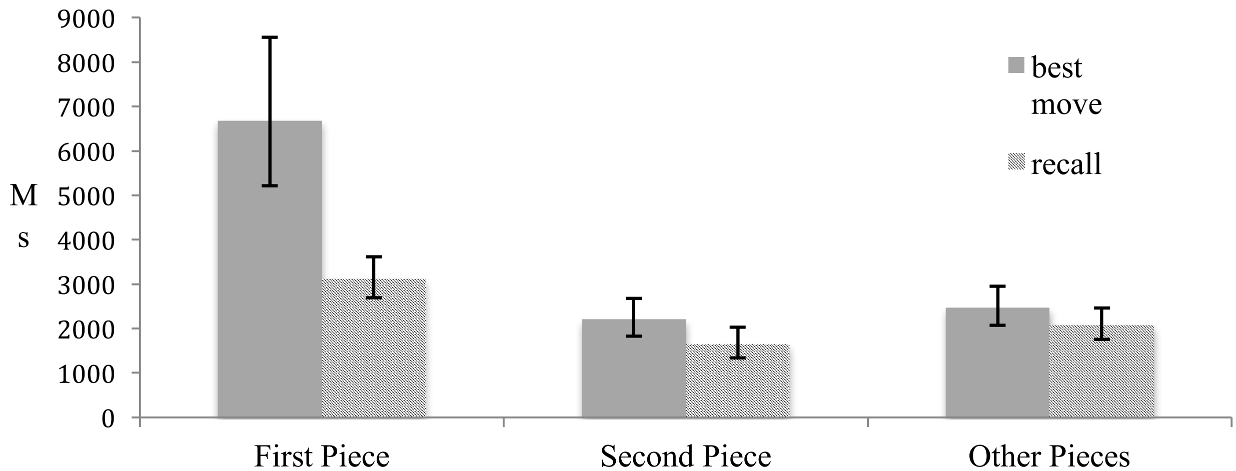

Figure 4. Means and 95% confidence interval for response times for various piece orders in both conditions. All means and confidence intervals were computed by retransforming the log response time variable.

was 3.39 (SD=.17) while for recall it was 3.31 (SD=.16). This translates to a difference of about 385 ms, This also was significant t(21)=4.06, p=.001, d=.87,

After the reaction times (RTs) for the first two pieces had been removed from the analysis, a repeated measures ANCOVA was conducted to test if the speed difference of the RTs for the two conditions differed by rating or age. In addition, an analysis was performed to test if the effects of condition were fairly normally distributed as would be expected if the condition differences in RT had errors around a mean difference. As would be expected based on the t-test above, the effect of condition was significant *F*(1,19)=20.68, p<.001. The interaction between chess rating and task condition was not significant *F*(1,19)=.22, p=.65, but the interaction of age and condition was significant *F*(1,19)=5.61, p=.03. The slowed RTs due to age appear to be greater in the best move condition β=.014, t=5.56, p=.001, than in the recall condition β=.010, t=3.05, p=.007. The between subject's analysis showed no effect of rating *F*(1, 19)=1.02, p=.32

the direction was for the better players to be faster, but as would be expected the older players were in general slower to respond $F(1,19)=11.66$, $p=.002$. Histograms for the best move condition's

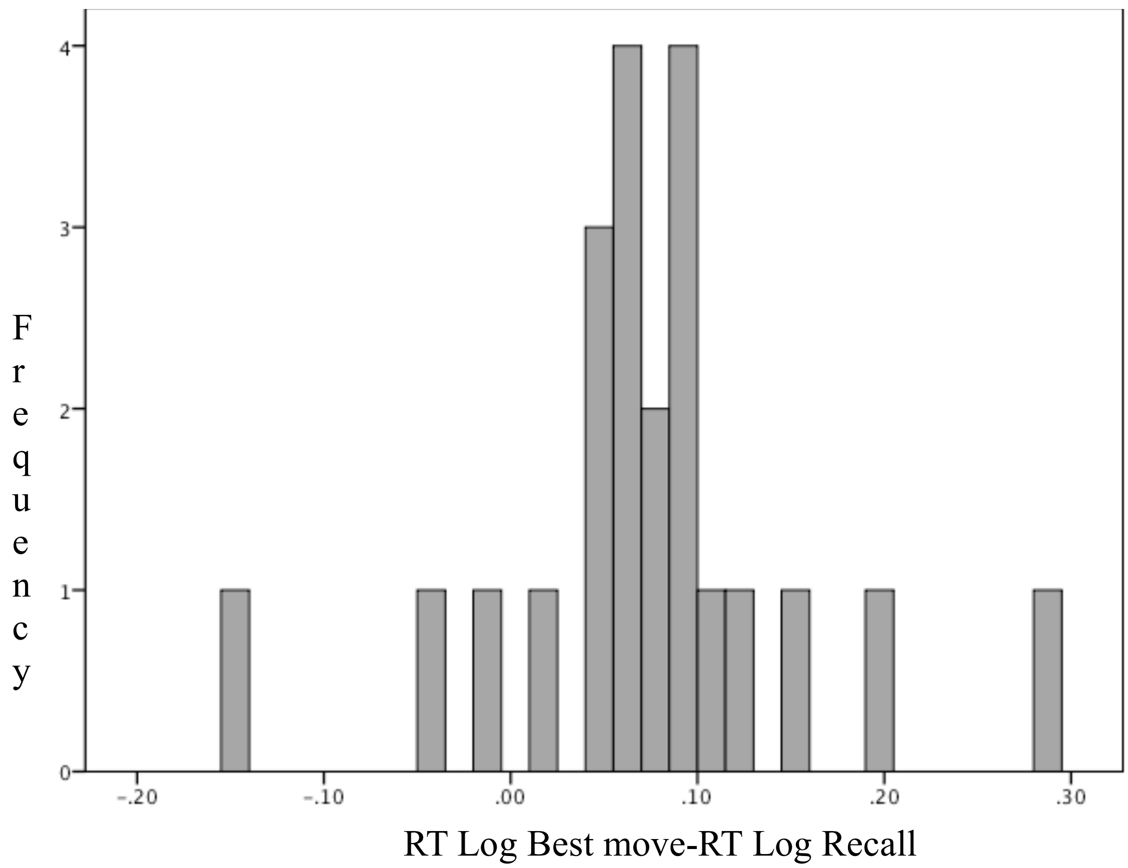

Figure 5. Histogram showing the distribution of the differences in the conditions response times for each participant.

RTs are shown in Figure 5. The distributions do not diverge from normality as per the Kolmogorov-Smirnov test=.17, $p=.12$. Additionally, while in general participants who got more correct did not display any greater average RT difference between the two conditions $r(23)=-.32$, $p=.15$. However as would be expected, the correlation with the squared distance from the grand mean for all participants in the RT difference between the two condition, which can function as a test of heterogeneity, is significantly correlated with the average number of pieces recalled correctly $r(23)=-.44$, $p=.004$. This suggests that participants producing more responses tended to have an RT difference between conditions closer to the average RT difference between conditions for all participants. These results support the suggestion that the effects were not confounded with overall performance and that as more responses were accumulated, leading to a more reliable estimate, the difference converged towards the grand mean.

The full distribution of RTs for recalling pieces other than the first two pieces are shown in Figure 6. The overall means are 3.28 (SD=.38) for the recall condition and 3.34 (SD=.41) for the best move condition and when untransformed this equates to a difference of roughly 300 milliseconds (299.36). The difference in RTs across percentiles of response speeds for both log transformed and untransformed are illustrated in Figure 7. As can be seen at every percentile the best move condition is slower however both transformed and particularly when untransformed show much bigger differences to appear to be at higher percentiles, which corresponds to slower responses. The log transformed data, once the mean shift is removed by condition centering though the distributions, are not significantly different from each other (Kolmogorov-Smirnov test=1.24 $p=.09$). Figure 8 shows the distribution of log RTs of the best performing participants. Interestingly no evidence of bimodality is seen in any of the distributions.

Additional Analysis of Data from Study 2

The primary finding from the analyses of memory of the chess positions in the two conditions concerned the significantly bigger largest chunk in the memory condition. It seems natural to ask, if the rules for determining the chunk boundaries were changed for the best move task to reflect the general slowing, and the greater slowing, for the recall of the second piece in each trial in the best move condition, would the inferences about the differences in bigger largest chunk remain. In order to further understand the differences in recall times for consecutive pieces in the two conditions, several analyses were conducted to examine if the recall in the best-move condition was influenced by the piece selected to be moved. In addition, the semantic relations between consecutive pieces were analyzed to examine potential differences between the two conditions.

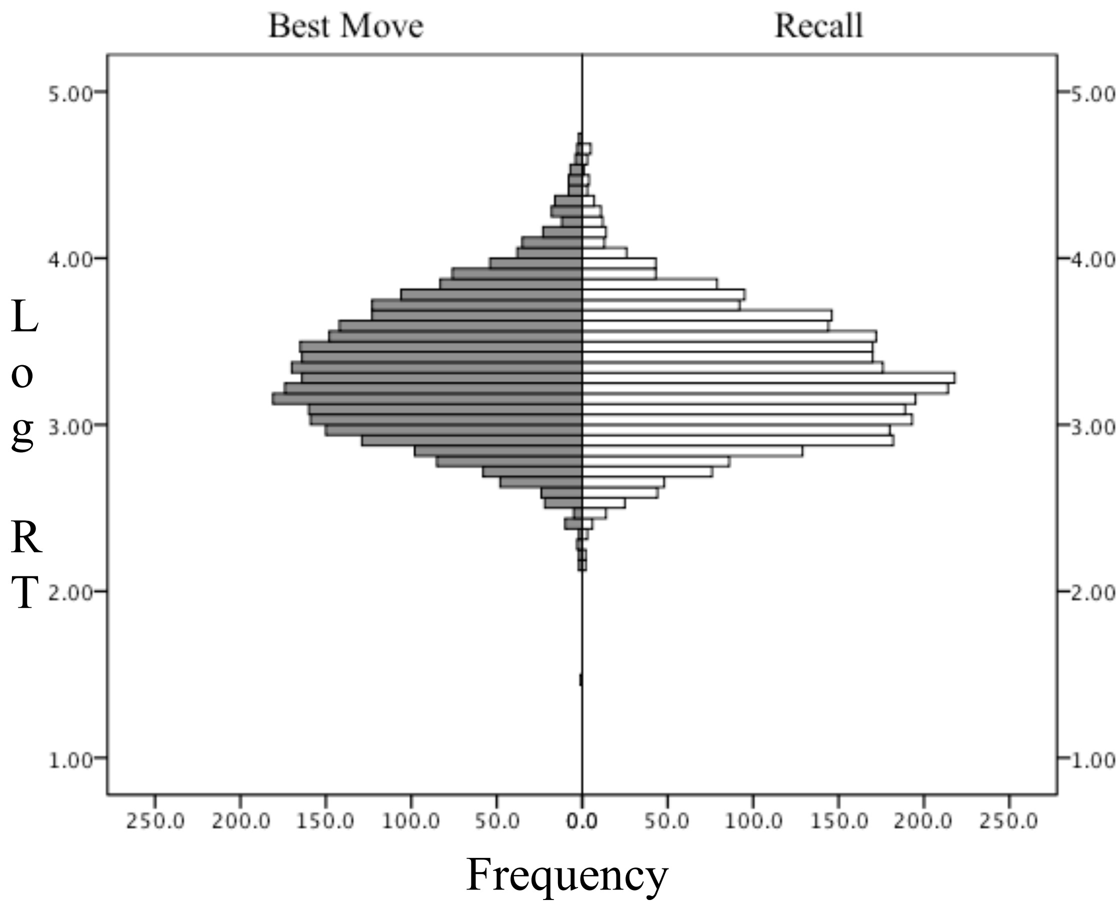

Figure 6. Population pyramid showing all responses times for each piece placed after the first two pieces for both conditions

Re-analysis of chunk variables. Therefore, the chunking variables of number of chunks and size of the largest chunk were recalculated correcting for the longer recall times in the best move condition. The difference of 300 milliseconds was used which was the overall difference

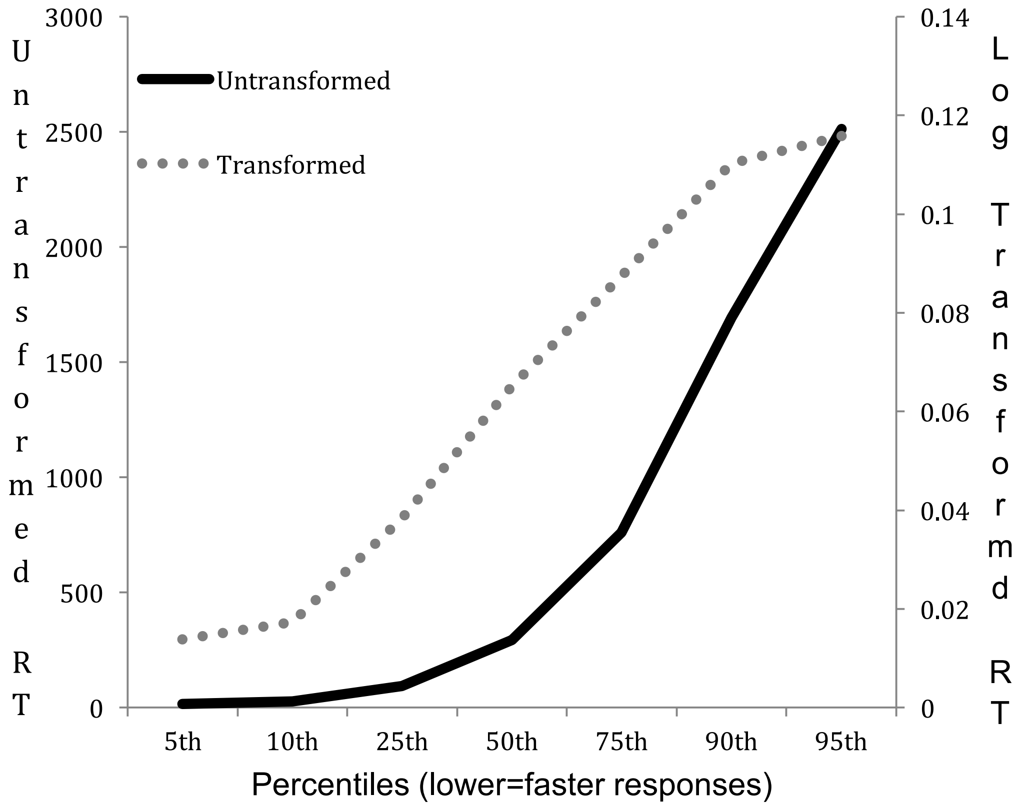

Figure 7. Shift function showing an increased slowing of RTs in the best move condition both for the Log transformed times (Transformed) and with those times untransformed (Untransformed) transformed RTs for the 3 best performing subjects.

between all responses, except for the second piece where 550 was used which was the equivalent difference between all responses.

After lengthening the RTs required for identifying a new chunk boundary during recall, the recall after the best-move task had an average of 6.73 (SD=3.51) chunks with the largest chunk being on average 4.36 (SD=2.72) pieces. A new analysis with the same multilevel models analyzed showed a similar absence of significant differences in the number of chunks t(506)=-.1.25, p=.21, and the size of the largest chunk was not significantly different t(508)=-1.87 p=.06.

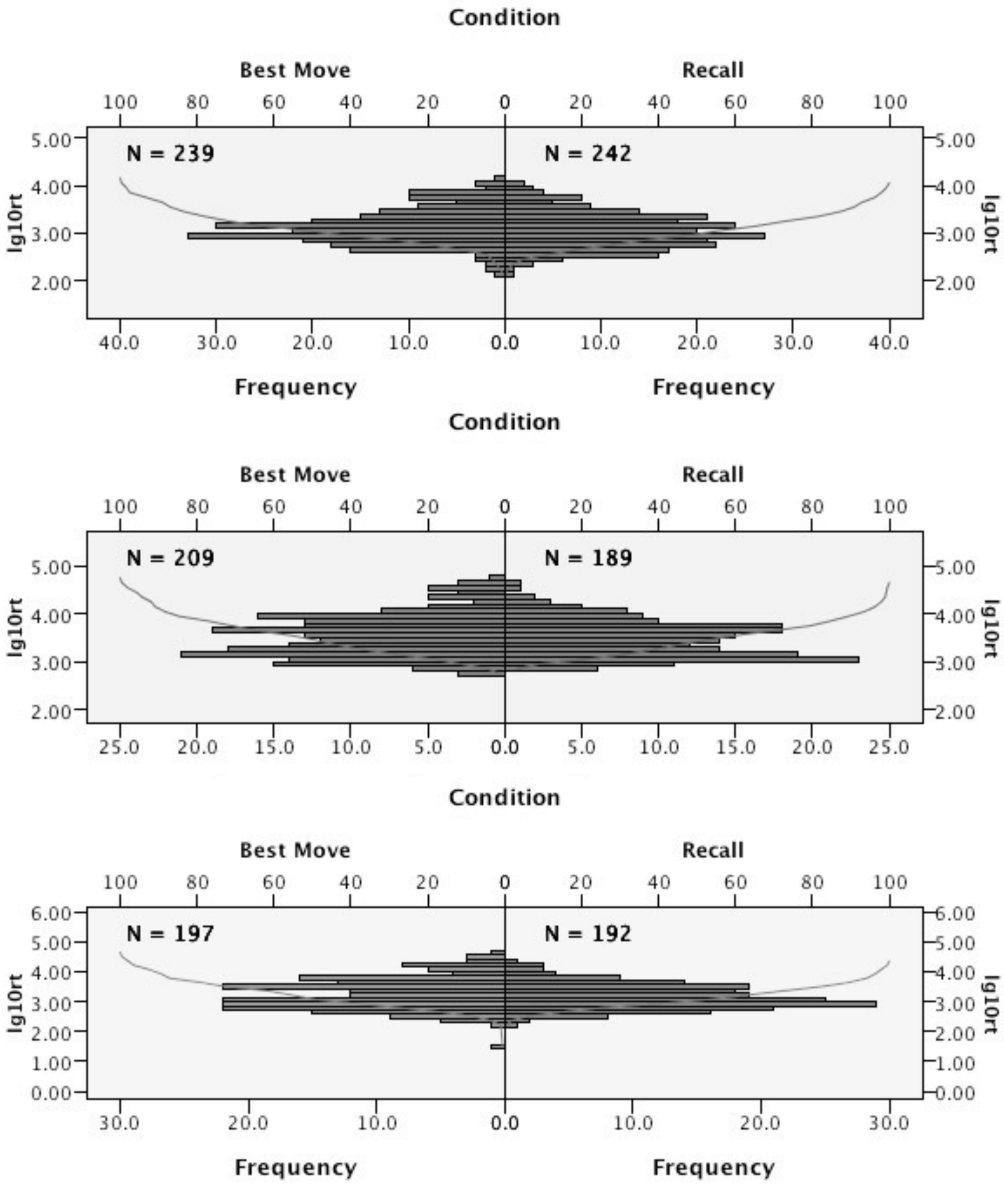

Figure 8: Histogram of log transformed RTs by condition for the best performing participants.

Features of the Selected Piece in the Recall after Selecting the Best Move

The next issue analyzed concerned the effect of the piece selected in the best-move condition on the recall of chess pieces in that condition. The most likely effect on the structure of recall is that the selected piece would be privileged during the subsequent recall. Another likely effect on the recall is that the associations between the selected piece and other pieces directly related to that piece would be more likely encoded and thus included in the subsequent recall.

Subsequent recall of the piece selected to be moved in the best-move task was significantly higher .68 (SD=.27) compared to .34 (SD=.18) for all other pieces in the presented chess position, $t(22)=7.60$, $p<.001$, $d=1.58$. When the piece selected to be moved was correctly recalled in its presented location it was recalled at the first part of the subsequent recall. The serial order of the chunk with the selected piece was on average 1.94 (SD-.83) while the average midrank in serial position recalled chunk was the 3.86 chunk recalled (SD=1.56, $t(22)=-5.23$, $p<.001$, $d=1.09$. This means that the selected piece was more often recalled among the first chunks recalled as opposed to one of the later chunks. The average order of recall for the selected move was 2.87 (SD=1.94) while the average mid-ranked recalled piece was the 6.74 piece (SD=3.14), $t(22)=-5.08$, $p<.001$, $d=1.06$ which means that the moved piece was typically one of the earliest pieces recalled. Since there are no ties in serial position the average serial order, median serial order, and midrank serial order all yield the same values, the midrank calculation was used because it was computationally easier to calculate for the chunk values. Additionally, errors could be included, therefore the number of pieces in no way should correspond to the number correct.

Analysis of Piece Relations between Consecutively Recalled Pieces

To examine if the selected piece to be moved was recalled in a similar manner to other pieces' recall in the recall condition a matched control piece from that condition was identified. The matching of pieces was done by using the same board across condition e.g. the sixth best move position presented was matched to the sixth recall position presented. Within that position the chunk number was matched or the piece number was matched using the same procedure as described in the previous paragraph. The types of relationships are attack (one of the pieces is attacking the other piece), defense (one of the pieces is defending the other), proximity (the pieces are next to each other), color (both are the same color), and piece (both are the same type of piece). The question that these analyses seek to answer is: was the piece unique in the chess information it contained, compared to a control?

For next piece relationships a proportion of .47 were on the chunk boundary which is defined by a piece being the last piece recalled in the chunk (SD=.50). Attack relations happened at a rate of .11 (SD=.31) compared to .40 (SD=.49) for defense relations. The most common relationship was color at .71 (SD=.45). The rates for proximity was .51 (SD=.50) and for same piece relationship it was .33 (SD=.47).

To predict the chess relationships between two consecutively recalled pieces a binomial multilevel model was fit with the predictor variable of if the piece was the boundary piece (yes=1 or no=0 item level variable), condition (item level), rating (subject level), and age (subject level). Board was not used as a random effect as it was too small, but subject was included. The LMER program in R was used (Kuznetsova, Brockhoff, & Christensen, 2013)). LMER reports estimates, z values, and p values. For attack relationships the only significant predictor was condition $\beta=.88$, $z=1.99$, $p=.04$ meaning there were more attacks by the selected piece than by the matched control piece. The results for being on the chunk boundary was $\beta=-.55$, $z=-1.25$, $p=.21$, for rating it was $\beta<.01$ $z=.61$,$p=.54$ and for age it was $\beta=.02$, $z=.63$, $p=.53$. For defense relations condition was again significant $\beta=-89$, $z=-3.22$, $p=.001$ this time with the matched piece in the recall condition defending the following piece more often. The estimates for being on the chunk boundary was $\beta=-.15$, $z=-.54$, $p=.58$, for rating it was $\beta<.01$, $z=1.78$, $p=.07$, and for age $\beta<-.01$, $z=-.23$, $p=.81$, For the proximity relationships the only significant predictor was being on the chunk boundary β=-.66, z=-2.37, p=.02, which means there were fewer proximity relationships on the chunk boundary. For condition the trend was for more relationships between pieces in the recall condition β=-.23, z=-.88, p=.38, for rating it was β<-.01, z=-.39, p=.70, and age β=-.01, z=-.44, p=.66. For the relationship of color there were no significant predictors. The estimates were β=-.55 z=-1.81, p=.07, for chunk boundary β=-.51, z=-1.75, p=.08, for condition, β<-.01, z=-.09, p=.93, for age and β<-.01, z=-.29, p=.78, The relationship of being the same piece occurred less frequently at the chunk boundary β=-.63, z=-2.07, p=.04, and in the best move condition β=-.63, z=-2.16, p=.03, For rating the estimate was β<.01, z=.30, p=.76, while for age it was β=-.03, z=-1.36, p=.18,

Additionally, an analysis was conducted to test if any of the included variables predicted if the piece was on the chunk boundary. No variable predicted being on the chunk boundary. The estimate for condition was β=.40, z=1.47, p=.14, for rating the estimate was β<.01, z=.71, p=.48, while for age it was β=.03, z=1.38, p=.17,

For total number of relationships (estimated in LMER with a Gaussian distribution assumed) as suggested by past research there were fewer relationships at the chunk boundary β=-.50, *t(*235)=-3.22, p<.01. Additionally, in the best move condition there were fewer relations on average β=-.42, *t(*235)=-2.77, p<.01. No other variable predicted total number of relations. For rating the estimate was β<.01, *t(*19)=.79, p=.44, while for age it was β=-.01, *t(*19)=-1.00, p=.33.

Analysis of Spatial Location of Recalled Pieces

The next analysis will investigate where on the chess board pieces were recalled, by condition. The eye movement patterns described by past research for the recall task have involved a circle search around the board (de Groot & Gobet, 1995) which is a very different pattern than the pattern described by Tichomirov and Poznyanskaya (1966) for move selection. It might therefore be the case that the location of the recalled pieces will vary by condition. For these analyses the board was divided into four sectors. The first is the center, defined here as e4, e5, d4, and d5 on the chessboard. The proportion of pieces recalled in this sector was .24 (SD=.19). Next was the set of squares outside the center sector which will be called center plus 1, the proportion of pieces recalled in this segment was .27 (SD=.17). Next was the set of squares separated from the center by the previous set. Those squares which will be called the center plus 2, participants recalled a proportion of .32 (SD=.20) pieces in this segment. Finally was the outer ring of the board containing the back ranks for white and black as well as the A column and the h column which will be called the outer squares. The proportion of pieces recalled in the outer ring was .35 (SD=.20). Figure 9 highlights the sectors that were coded.

A 2X4 repeated measures ANCOVA was conducted using condition and the 4 types of squares controlling for age and rating. Both rating and age were centered and degrees of freedom are Greenhouse-Geiser. There was no effect of condition $F(1, 20)=.14$, $p=.71$, $\eta^2_{(partial)}=.01$, nor interaction of condition and rating $F(1, 20)=.92$, $p=.35$, $\eta^2_{(partial)}=.04$, nor interaction of condition and age $F(1, 20)=.37$, $p=.55$, $\eta^2_{(partial)}=.02$. There was a main effect of type of square $F(1,33.67)=.11.81$, $p<.001$, $\eta^2_{(partial)}=.37$. This effect did not interact with rating $F(1,33.67)=.37$, $p=.66$, $\eta^2_{(partial)}=.02$ or age $F(1,33.67)=.75$, $p=.46$, $\eta^2_{(partial)}=.04$. Condition and type of square did not interact $F(1,32.64)=.69$, $p=.56$, $\eta^2_{(partial)}=.03$. The three-way interaction with rating was not significant $F(1,32.64)=.41$, $p=.75$, $\eta^2_{(partial)}=.02$, nor was the three-way interaction with age $F(1,32.64)=.34$, $p=.67$, $\eta^2_{(partial)}=.02$. Probing the effect of type with a Bonferoni corrected p value of .0083 showed that the outer squares were recalled more often than the center

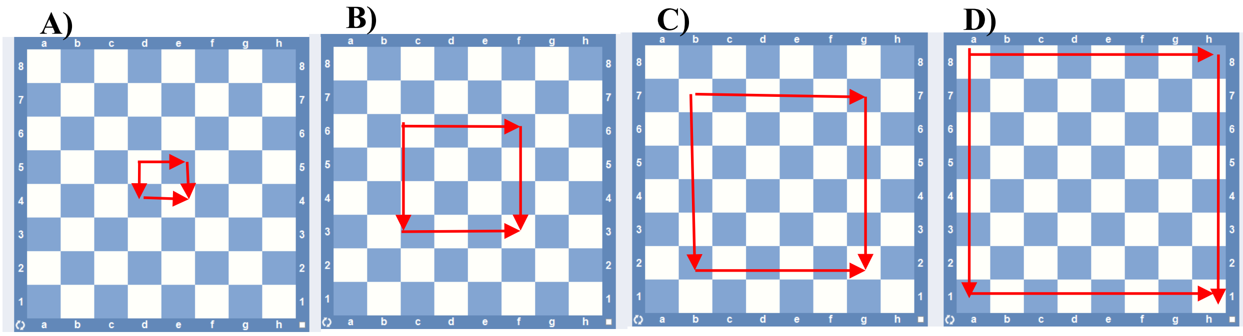

Figure 9: Figure of coding of board sectors, A) highlights the center, B) highlights the center plus 1 squares, C) shows the center plus 2 squares, and D) highlights the outer squares.

$t(21)=3.71$, $p=.001$, $d=.82$. The outer squares were also recalled more than the center plus 1 squares, $t(21)=4.00$, $p=.001$, $d=.85$, but not the center plus 2 squares $t(21)=2.48$, $p=.02$, $d=.53$. The center plus 2 squares were recalled more than the center $t(21)=3.33$, $p=.003$, $d=.71$ and the center plus 1 squares $t(21)=3.19$, $p=.004$, $d=.68$.The center and the center plus 1 ring did not significantly differ $t(21)=1.46$, $p=.16$, $d=.31$.

Discussion Study 2

Study 2 mainly tested a set of null hypotheses designed to indicate if the structure of chess memory would be the same if the goal of the primary task was manipulated. While for the number of items correct there were no differences, for the structural features such as size of the largest chunk and number of chunks there were clear differences. The most substantial difference appears to be a fairly large RT difference which may have caused most of the other differences. This was supported by the reanalysis which found after controlling for RT differences the chunking variables were not significantly different. The findings from the current study largely fit with past research. First even among this restricted age range, age was negatively related to performance and was associated with smaller chunks as shown by past research (Charness, 1981). Similarly, as is important for chunking and Template theory, expertise was related to larger chunks, but not more chunks.

There are additional checks that were done to test how similar this data is to previous data. Given that there were no interactions with condition, all these will be reported for the full data set. Moxley & Charness (2013) estimated a relationship of rating and recall performance of .51 95% CI[.42, .59] and age and recall of -.49 95% CI[-.36, -.60]. The partial correlations in this study are well within those ranges at .56 for rating and -.44 for age. Moxley & Charness (2013) estimated a relationship of rating and move selection of .76 95% CI[.64, 83] and age and move selection of -.27 95% CI[-.18, -.37] but the partial correlations in this study are outside those ranges at .46 for best move and -.41 for age. This means that for a dual task and rapid move selection rating was less positively related to move choice than would be expected by previous studies and age was more related. Finally, on the overall results Gobet and Simon (1996a) class A players while rated slightly higher on average than our mean had similar performance recalling 34.5% of the pieces correctly compared to 34% in this study. Gobet and Simon (1998) also found a similar number of 32% seemingly well within the expected range.

The investigation of next piece relationships confirmed that the moved piece was more likely to have attack relations with following piece and less likely to have some other relations (defense and being the same piece with the following piece). The moved piece was also special in several ways, being recalled earlier and generally being from a larger chunk. The effect of it being from a larger chunk than the control chunk is surprising because in general the largest chunk from the best move condition was smaller than from the recall condition if RT was not controlled. This all suggests that selecting a move changes recall, although caution has to be exercised as the analyses were very exploratory.

Finally, it is surprising that between subjects the strength of move selection was a better predictor than rating. This is probably in part due to participants who were struggling on one aspect of the task and also struggled on other aspects of the task. More analysis is needed on this point. Preliminary analysis of the recall condition supports the hypothesis that participants were either generally doing well or generally struggling as move strength predicts number correct, but did not show a similar effect for size of the largest chunk. After controlling for rating, it also did not appear that the relationship between move strength and recall was caused by problems with a specific board as move score did not predict significantly at the item level. Instead it appears players who were able to rapidly select an acceptable move were also able to correctly recall a larger number of pieces in general.

Differences in the structure of recall means that Study 2 failed to support the key assumptions of various pattern recognition theories. The knowledge structures of chess (at least as measured by chunks) do not appear to be activated automatically by exposure to a given chess position. This is not necessarily surprising as Tichomirov and Poznyanskaya (1966) famously showed that when searching for a move chess players' eye movements tended to follow lines of relationships between pieces particularly in attack and defense relationships. DeGroot and Gobet (1995) eye tracking for recall showed that players scanned around the board in a circular pattern. That kind of difference would not necessarily mean that the knowledge activated would be different, but it does make it less likely. Based on this difference in the literature n exploratory analysis tried to test if the probability of a piece being recalled varied based on the location of the piece, but yielded a null finding.

GENERAL DISCUSSION

This study makes a number of important contributions. The first is that it showed that chess skill could be reliably measured with the presentation times used for the administration of recall tasks. Conducting move selection experiments with brief presentation times offers new opportunities for lines of research and manipulations/process tracing that might not be feasible with longer times. The second contribution is this study extended past research on the relationship between speed and skill in chess to an experimental task. This study supports past research by Moxley et al. (2012) that the effects of skill do not vary generally with speed, but only in specific situations. The final contribution was that recall of presented pieces was not invariant across the conditions. While the mean number of pieces recalled correctly did not vary between conditions, there were strong differences in the chunking effects likely caused by a large general slowing of response times in the best move condition. This result argues against the theory of automatic activation of knowledge representations regardless of task

In the introduction of this paper I laid out three goals: first to better unify the evidence from the best move with the recall task. Study 1 confirmed that 5 seconds was sufficient to get reliable skill differences, and Study 2 showed that it is possible to request recall after the completion of move selection while using the same types of problems. The second goal was to test assumptions of pattern recognition theories: that expertise was defined by rapid processes and that the knowledge structures of chess can be described by chunks automatically accessed. The third goal was to update our theoretical understanding of the knowledge that supports chess decision making. In the following section I will discuss progress with the last two goals.

Implications for Automatic Pattern Recognition Theories

Cost of Time Pressure in Relation to Skill

Leading theories of expert intuition (Klein, 1998) suggest that one of the defining features of expertise is to have an increasing advantage as time demands increase. Theories of expert pattern recognition have accepted the findings that lead to these claims and argue that fast processes are far more important for high level skill than are slow processes (Gobet & Simon, 1996b). A recent paper published in Journal of Experimental Psychology: General summarized our knowledge this way: “Expert players excel specifically at rapid object recognition abilities and, hence, under temporal pressure are expected to further amplify the differences with weaker players.” (Fernandez, Slezak & Sigman, 2012, p. 528). The evidence for this claim has been more anecdotal than has been acknowledged up to this point. There are three main studies cited as evidence for the conclusion of an increased expert advantage in rapid chess play. The first is Gobet and Simon (1996c), which examined Gary Kasparov’s performance in games where he would have had 30 seconds per move compared to his more normal 3 minutes per move. Gobet and Simon (1996c) found a seemingly small decrease in performance. This study does give an example of how much skill can be maintained by an elite player when put under time pressure. However, the claim that experts are relatively more reliant on intuition would require establishing that Kasparov was unique in this lack of a decrease in performance. Gobet and Simon (1996c) did not attempt to establish this. Nor could they control for other factors such as Kasparov’s likely much greater rote knowledge of openings as well as the quality of his preparation for each opponent compared to their preparation for him. Burns (2004) showed less underperformance of their chess rating for higher rated players in quick tournament games for games played against relatively lower rated players than for weaker players facing lower rated players. However, this design does not permit testing an interaction of rating and available time. The method Burns (2004) used was not validated for potential biases so it is impossible to know how resilient this measure would be in measuring well known effects such as the increased reliability of performance at higher levels of skill (Glickman & Jones, 1999). The issue that if performance is more variable at lower levels of skill that might be expressed by weaker players under performing their rating against other relatively weaker players and over performing against relatively stronger players. Since Burns (2004) only looked at one side this potential confounds cannot be ruled out.

The most compelling study, and the only one to test the interaction and find it, was Calderwood et al. (1988). This study however, has several problems. They measured performance for players by analyzing the average ratings of moves in entire games in a group of games between two experts compared to games between two much weaker players. It is not clear how well this system would perform in distinguishing skill in a single game. A main concern is that the only reason the study found the interaction is because there was no difference between the experts and the non-experts in classical chess; this difference is one they should have found due to the huge skill differences between the groups.

As noted in the discussion of Study 1, the three-way interaction does open the possibility that those studies are correct, but this study only supports that conclusion if most chess positions are relatively easy and have a clear best move. In challenging situations experts benefited relatively more from extra time. Over the entire 50 question test given in this study drawn from very heterogeneous games (the skill range of the players who made the move the positions were drawn from ranged from 1400-2600) there was no evidence of a general interaction of skill and time. Additionally, Calderwood et al.'s (1988) study suffered from potential bias causes by judging every move of a full games and having a grandmaster judge the games. It is impossible to know what idiosyncrasies were in the games as well as in the scoring. The current study used tight experimental control with a much larger sample than was used by Calderwood et al. (1988). From personal experience of computer analyzing thousands of chess games I will say one chess game is simply not a reliable proxy for skill (see the lack of correlation in blunders between two games by the same players described in Jeremic, et al. (2010). The best move task also tends to correlate with skill almost as well as a full tournament (Moxley & Charness, 2013), which obviously is a better predictor than a single game. Finally, computer analysis as used in this study should be less prone to bias than master player ratings as should the use of problems designed to allow discrimination across skill levels.

The findings of Burns (2004) have already been called into question by Van Harreveld et al. (2007) finding the opposite pattern of the expected finding, e.g. they found a lower correlation at faster speed between online performance and tournament rating. Van Harreveld et al.'s (2007) study is unfortunately also flawed as it does not validate how well online ratings should correlate with skill within restricted ranges. The biggest concern is some chess players play literally hundreds of thousands of games online (see Ericsson & Moxley, 2012, for an example) and it is not clear how playing that many games, most of which are against hopelessly overmatched players, will affect rating. Evidence for inflation of some players' ratings in the Van Harreveld et al.'s (2007) study is given by the fact that some of the players in this study have ratings over 3000, a level exceeding even the World Champion, Magnus Carlsen's record level by over 200 points. The strongest methods and strongest analysis ever used to test this hypothesis were used by Moxley et al. (2012) and in the current study. Neither came close to identifying even a meaningful general trend in either direction. At this point it simply cannot be claimed that we know that chess skill differences increase as time decreases or to claim the opposite.

The question of a general interaction between skill and time should be viewed as an open question and one warranting even further experimental studies. Currently the evidence better fits only very specific situations as having this effect, balanced by others having the opposite effect. A potential follow up would be to take a broader, possibly random, sample of chess positions. The 30 and 5 s conditions appear to be sufficient to test this hypothesis and with only those two conditions it would be possible to do as many as 100 problems in an hour, which should allow a sensitive test of the hypothesis.

Implications for Theories of Chess Knowledge

Problems with chunking theory were raised already in the mid 70's (Charness, 1976) when it was demonstrated that STM processes could not by themselves explain chess memory. Template Theory addressed these issues by adding templates to the model which could both be larger and had slots that acted as retrieval structures allowing more rapid and reliable encoding into long-term memory. The principle difference between template theory and chunking theory concerns the addition of templates and templates' ability to rapidly store information in long-term memory. This study was designed to test assumptions of automatic pattern recognition models and unfortunately has demonstrated several new examples of issues with chunks as identified from the recall task.

One of the first problems is with the idea of the chunks as a unit of semantic knowledge acquired through skill. One piece of evidence for chunks as identified by the recall task as being a unit of knowledge is that next piece relationships between consecutive pieces within chunks exceeds the number of relations between consecutively recalled pieces belonging to different chunks (between chunks). This study replicated this difference when the relations to the subsequent piece recalled immediately after the moved piece and a matched control piece from the recall condition. It should, however, be noted that it was only those relationships between chess pieces that would be apparent to a non-player of chess that showed this effect. Attack and defense relationships did not show any difference between chunks. Linhares and Chada (2012) noted that piece relationships do not distinguish skilled from unskilled players in many studies. Gong et al. (2015) showed that at least semantic relationships (defined as attack or defense relationships) did distinguish players from non-players. The current study went a step further than previous studies by testing each type of relationship for differences between and within chunks. Study 2 found that it is the type of relationships that would be obvious to a non-player that can be discriminated by testing if the next piece is within or between chunks. Consistent with Linhares and Chada's (2012) concerns, this raises the question of how can these chunks define skill if they are themselves defined by features obvious to someone who has never played the game?

The assumption that the structure of the patterns recalled reflect recall of patterns retrieved automatically is inconsistent with that based on five sets of evidence. First there was a general slowing in the best move condition suggesting that a very different process of recall was being engaged. Secondly, the chunking variables identified from the recall suggested bigger largest chunks, though this may be a consequence of the slower recall times. Thirdly, the moved piece appears to be privileged in recall and is more likely part of a larger chunk with different piece relationships compared to matched control pieces from the recall condition. Again this suggests that the task condition causes some changes in recall, although these analyses were exploratory. We do not know what piece, if any, they would have selected in the recall condition if they had had a chance to select a move. Fourthly, the general slowing of recall in the best-move condition is inconsistent with an argument for automaticity and instead suggests a controlled process of recall that is different when done after move selection. And finally the RT distributions do not appear to be noticeably bimodal as would be expected by chunking theory. Far from being automatic it appears that engaging in problem solving causes either an additional process or a different process to be engaged during the retrieval of pieces. This processing requirement does not appear to be deleterious to performance, but is pervasive throughout the recall process.

While the transformed distributions are not likely normally distributed no bimodality is clear by visual inspection. While some might argue that RT distributions come from ex-Gaussian distributions and perhaps the exponential component, which would be removed by the transformation, is caused by assessing a different chunk in memory. This is a very strong assumption and, at least so far, research has not supported that two cognitive processes can be inferred from ex-Gaussian distributions (Matzke & Wagenmakers, 2009).

The Knowledge behind Decision-Making in Chess

The strong suggestion of this research is that chunks as derived from the recall task are simply not useful to understanding the expert knowledge structures. There are simply too many issues with them at this point. Three particular issues were demonstrated by this study, first that they are not invariant across task conditions as they should have been if they were fundamental units automatically accessed in response to the exposure to a chess position. The second was that the features that differentiate sequences of pieces recalled within chunks from consecutively recalled pieces categorized as belonging to different chunks are salient to a non-player and do not reflect a higher-level understanding of chess. And finally, the distribution of response times does not appear to show strong evidence of bimodality, as it should.

Chunks could still be the unit of information that defines chess expertise, but it may not be possible to identify them through the recall task. It is possible that the copy task, which uses glances between periods of recall, or the circle chunks task, whereby players are asked to circle groups of pieces they believe are related, might better represent chess knowledge. One obvious advantage of chunks is that they make certain that players recall the whole position without errors. While it is possible that, particularly in brief situations, the representation of the position is only partial and is distorted, we can't know how much information is lost as recall is continued. The issue is obviously the copy task had been traditionally untimed so in theory the knowledge representation could change as the participant views the stimulus. One future method to test if these two task are identifying some stable representation would be to have players perform both task on the same positions at different times. If the chunks are basic unit of information, then the chunks should overlap heavily and have very consistent features between the two tasks. Additionally, on the copy task the directions for the order of copying the position could be varied. For instance, chess players complain during the recall task about having to choose each individual piece, preferring to place all the pawns at once. If the chunks are stable units of memory it should not matter the modality of recall (one piece selected at a time compared to being able to put down multiple pawns for instance) when asked to copy the position. It strikes me as unlikely this would be the case. Additionally, both task could be done using the same primary task manipulation used in this experiment.

It is not clear at this point what exactly the knowledge structures behind chess move selection are, but we know some of their features. The first feature is that it is apparently flexible.

For instance, experts can synthesize chess pieces presented individually in random order into a meaningful enough representation to allow move selection as accurately as if they were presented with the whole position at the same time (Saariluoma & Kalakoski, 1998). It apparently can be invoked incidentally with some slowing when selecting moves. Finally, it allows experts to alter how they represent positions based on instructions, something non-experts are not always able to do. (Linhares & Blum, 2009).

Limitations

These studies have a few limitations. First the studies instructed participants to "think aloud", even with brief times for generating their responses. Eliciting "think aloud" protocols do not typically interfere with problem-solving performance but these verbalization processes can slow down the generation of responses (Fox, Ericsson, & Best, 2011). This raises the potential concern that at least in the 5 s conditions some participants may have performed better had they not been instructed to "think aloud". This potential effect may account for the larger than normal age effects for move selection in the 5 s condition used in study 2. A second issues concerns the criteria for selecting move-selection problems for the studies. The criteria used in this study have many strengths, however, the results of this study may not unconditionally generalize to move-selection with other problem sets selected with different criteria. Finally, while the analysis of move strength was consistent for both the Winsorized move score and the correct or incorrect scoring the three-way interaction with move-strength was not tested using both methods of scoring. Additionally, while the method of scoring the moves had clear strengths and was based on extensive experience with analyzing chess games and chess moves there are other methods of scoring move strength such as player ratings or simply using the raw score.

Conclusion

This paper has called into question two of the most basic things researchers thought we knew about chess. After this work it should not be said that we know experts have an extra advantage in fast time or that the chess knowledge structures are automatically accessed. Instead chess skill maybe stable across time but flexible to the needs of the task. Both of these are consistent with the idea that experts gain greater understanding of a position from study and that this representation is tuned to the needs of the situation (Ericsson & Kintsch, 1995). If this is the case all the expert advantages gained from analysis of a position may not show up in the initial position but instead would be played out over multiple positions.

APPENDIX A

IRB APPROVAL

The Florida State University

Office of the Vice President For Research

Human Subjects Committee

Tallahassee, Florida 32306-2742

(850) 644-8673, FAX (850) 644-4392

RE-APPROVAL MEMORANDUM

Date: 3/25/2015

To:

Address:

Dept.: PSYCHOLOGY DEPARTMENT

From: Thomas L. Jacobson, Chair

Re: Re-approval of Use of Human subjects in Research

Speed knowledge and aging in chess

Your request to continue the research project listed above involving human subjects has been approved by the Human Subjects Committee. If your project has not been completed by 3/23/2016, you must request renewed approval by the Committee.

If you submitted a proposed consent form with your renewal request, the approved stamped consent form is attached to this re-approval notice. Only the stamped version of the consent form may be used in recruiting of research subjects. You are reminded that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report in writing, any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor are reminded of their responsibility for being informed concerning research projects involving human subjects in their department. They are advised to review the protocols as often as necessary to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

Cc: K Ericsson, Advisor [

HSC No. 2015.15122

The Florida State University

Office of the Vice President For Research

Human Subjects Committee

Tallahassee, Florida 32306-2742

(850) 644-8673, FAX (850) 644-4392

APPROVAL MEMORANDUM

Date: 5/23/2014

To:

Address:

Dept.: PSYCHOLOGY DEPARTMENT

From: Thomas L. Jacobson, Chair

Re: Use of Human Subjects in Research

Speed knowledge and aging in chess

The application that you submitted to this office in regard to the use of human subjects in the proposal referenced above have been reviewed by the Secretary, the Chair, and one member of the Human Subjects Committee. Your project is determined to be Expedited per 45 CFR §

46.110(7) and has been approved by an expedited review process.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals, which may be required.

If you submitted a proposed consent form with your application, the approved stamped consent form is attached to this approval notice. Only the stamped version of the consent form may be used in recruiting research subjects.

If the project has not been completed by 5/22/2015 you must request a renewal of approval for continuation of the project. As a courtesy, a renewal notice will be sent to you prior to your expiration date; however, it is your responsibility as the Principal Investigator to timely request renewal of your approval from the Committee.

You are advised that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report, in writing any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department, and should review protocols as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Human Research Protection. The Assurance Number is FWA00000168/IRB number IRB00000446.

Cc: K Ericsson, Advisor

HSC No. 2014.12717

APPENDIX B

CONSENTS

FSU Behavioral Consent Form

Factors determining acquisition of chess skill

You are invited to be in a research study on chess. You were selected as a possible participant because we want to study chess players. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by Jerad Moxley a research assistant in the cognitive psychology department.

Background Information:

The purpose of this study is to better understand the relationship between move selection and memory.

Procedures:

We will show you some chess boards and ask to select the best move from them. You will have different intervals to perform. While you do this task we will ask you to think out loud and we will tape that but we won't save any pictures of you. On a selection of the boards we will ask for a retrospective report after you have selected the best move. This is where you say again what you thought while solving the problem. We are asking for your name so we can confirm your chess rating. No one but me will be able to see your name. If you agree to be in this study, we will ask you a few questions about how you study chess. This will take about 90 minutes.

Risks and benefits of being in the Study:

The study has minimal risk to you outside of the cost of the time it takes. You will receive no benefits from participation outside of the payment made.

Compensation:

Chess players with USCF ratings or FIDE rating will be financially compensated for their time. If you are a student at FSU you will be compensated with course credit. Compensation will be $15 dollars an hour for players rated under 1800, $20 dollars an hour for players over 1800 and $20 dollars or tutoring rates up to $40 per hour for Master level players or above. You can withdraw for the study at anytime and you will be paid for the time you have spent (rounded up to the next hour). We will access ratings based off the current USCF list.

Confidentiality:

The records of this study will be kept private and confidential to the extent permitted by law. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. After being typed recordings will be erased. The survey and the demographics will be kept on opposite sheets of paper and coded before being studied. Your name and chess ID will be marked off the survey after coding and password protected in a data file only the researcher will have access to. The data file will be erased in 18 months. It will not be possible to figure out your answers after coding. Surveys will be kept securely for three (3) years after this study ends in a locked cabinet and office.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University or the US chess federation If you decide to participants, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions:

The researcher conducting this study is Jerad Moxley. You may ask any question you have now. If you have a question later, you are encouraged to contact him at (850)644-9850, or mo. The faculty advisor for this project is Anders Ericsson who may be reached at or.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), you are encouraged to contact the FSU IRB at 2010 Levy Street, Research Building B, Suite 276, Tallahassee, FL 32306-2742, or, or by email at humansubjects@magnet.fsu.edu.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

___________________ ____________________

Signature Date

___________________ ____________________

Signature of Investigator Date

FSU Behavioral Consent Form

Factors determining acquisition of chess skill

You are invited to be in a research study on chess. You were selected as a possible participant because we want to study chess players. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by Jerad Moxley a research assistant in the cognitive psychology department.

Background Information:

The purpose of this study is to better understand the relationship between move selection and memory.

Procedures:

We will show you some chess boards and ask you to learn them and then to reproduce them. Occasionally you will be asked to select the best move for the position before you recall it. We are asking for your name so we can confirm your chess rating. No one but me will be able to see your name. If you agree to be in this study, we will ask you a few questions about how you study chess. This will take about 90 minutes total.

Risks and benefits of being in the Study:

The study has minimal risk to you outside of the cost of the time it takes. You will receive no benefits from participation outside of the payment made.

Compensation:

Chess players with USCF ratings or FIDE rating will be financially compensated for their time. If you are a student at FSU you will be compensated with course credit. Compensation will be $15 dollars an hour for players rated under 1800, $20 dollars an hour for players over 1800 and $20 dollars or tutoring rates up to $40 per hour for Master level players or above. You can withdraw for the study at anytime and you will be paid for the time you have spent (rounded up to the next hour). We will access ratings based off the current USCF list.

Confidentiality:

The records of this study will be kept private and confidential to the extent permitted by law. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. After being typed recordings will be erased. The survey and the demographics will be kept on opposite sheets of paper and coded before being studied. Your name and chess ID will be marked off the survey after coding and password protected in a data file only the researcher will have access to. The data file will be erased in 18 months. It will not be possible to figure out your answers after coding. Surveys will be kept securely for three (3) years after this study ends in a locked cabinet and office.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University or the US chess federation If you decide to participants, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions:

The researcher conducting this study is Jerad Moxley. You may ask any question you have now. If you have a question later, you are encouraged to contact him at, or. The faculty advisor for this project is Anders Ericsson who may be reached at , or.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), you are encouraged to contact the FSU IRB at 2010 Levy Street, Research Building B, Suite 276, Tallahassee, FL 32306-2742, or 850-644-8633, or by email at humansubjects@magnet.fsu.edu.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

_________________ __________________

Signature Date

______________________ ______________________

Signature of Investigator Date

REFERENCES

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

Calderwood, R., Klein, G. A., & Crandall, B. W. (1988). Time pressure, skill, and move quality in chess. *American Journal of Psychology, 101,* 481-493.

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

Campitelli, G., Gobet, F., Head, K., Buckley, M., & Parker, A. (2007). Brain localization of memory chunks in chessplayer. *International Journal of Neuroscience. 117,* 1641-1659.

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 Performance and Perception, 7,* 467-47.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. Bower (Ed.) *The Psychology of Learning and Motivation* (Vol. 16, pp 1-58(. 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* (pp. 215-281). New York: Academic Press.

Chassy, P., & Gobet, F. (2011a). A hypothesis about the biological basics of expert intuition. *Review of General Psychology, 15,* 198-212.

Chassy , P., & Gobet, F. (2011b) Measuring chess experts' single-use sequence knowledge: An archival study of departure from 'theoretical openings' PLoS ONE, 6, e26692. doi:10.1371/journal.pone.0026692

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

deGroot, A. D. (1946/1978) *Though and choice in chess.* The Hague, The Netherlands: Mouton Publishers.

de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess.* Assen: Van Gorcum.

Dreyfus, H., & Dreyfus, S. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* New York: Free Press.

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

Ericsson, K. A., & Kirk, E. P. (2000). *Instructions for giving concurrent and retrospective verbal reports*. <http://supp.apa.org/psycarticles/supplemental/a0021663/a0021663_supp.html>.

Ericsson, K. A., & Moxley, J. H. (2012). A critique of Howard's argument for innate limits in chess performance or why we need an account based on acquired skill and deliberate practice. *Applied Cognitive Psychology, 26(4)*, 649–653. doi:10.1002/acp.2841

Ericsson, K. A., & Moxley, J. H. (2013). Experts' superior memory: From accumulation of chunks to building memory skills that mediate improved performance and learning . In T. J. Perfect & D. S. Lindsay. *The Sage Handbook of Applied Memory.* (pp. 404-420). New York: Sage.

Fernandez Slezak, D., & Sigman, M. (2012). Do not fear your opponent: Suboptimal changes of a prevention strategy when facing stronger opponents. *Journal of Experimental Psychology: General*, *141(3),* 527-538.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*, 316-344.

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

Glickman, M. E. & Jones, A. C. (1999). Rating the chess rating system. *Chance, 12,* 21-28.

Gobet, F., & Charness, N. (2006). Expertise in chess. In K.A Ericsson, N. Charness, P. J. Felrovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and Expert performance* (pp 523-538).New York: Cambridge University Press.

Gobet, F., & Chassy, P., (2008). Expertise and intuition: A tale of three theories. *Minds and Machines, 19,* 151-180.

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

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

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

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

Gobet F., & Simon, H. A. (1998b). Pattern recognition makes search possible: Comments on Holding (1992). *Psychological Research, 61,* 204-208.

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, and Cognition, 29,* 1082-1094.

Gong Y, Ericsson K. A., & Moxley J. H. (2015) Recall of Briefly Presented Chess Positions and Its Relation to Chess Skill. PLoS ONE 10(3): e0118756. doi:10.1371/journal.pone.0118756

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

Jeremic, V., Vukmirovic, D., & Radojicic, Z. (2010). Does playing blindfold chess reduce the quality of game: Comments on Chabris and Hearst (2003). *Cognitive Science, 34*, 1-9.

Kahneman, D. & Frederick, S. (2005) A model of heuristic judgement. In K. Holyoak & R.G. Morrison (Eds.), *Thinking and Reasoning* (pp. 267-293). New York: Cambridge University Press.

Kahneman, D. & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64,* 515-526.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95, 163-182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York, NY: Cambridge University Press.

Klein, G.A (1998). Sources of power: how people make decisions. Cambridge, Ma: MIT Press.

Klein, G. A., & Peio, K. J. (1989). Use of a prediction paradigm to evaluate proficient decision making. *The American Journal of Psychology, 103,* 321-331.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R.H. B., (2013) *lmerTest for random and fixed effects for linear mixed effects models (lmer objects of lme4 package):* R package version 3.2.2. http://CRAN.R-project.org/package-lmerTest

Linhares, A., & Brum, P. (2009). How can experts see the invisible? Reply to Bilalic and Gobet, *Cognitive Science, 33,* 748-751.

Linhares, A., & Chada, D. M. (2012). What is the nature of the mind's pattern-recognition process? *New Ideas in Psychology*, 31, *108-121.*

Linhares, A., Freitas, A. E. T. A., Mendes, A., & Silva, J. S. (2012) Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research, 13,* 72-86.

Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16,* 798-817.

Moxley, J. H., Ericsson, K. A., Charness, N., & Krampe, R. T. (2012). The role of intuition and deliberative thinking in experts' superior tactical decision-making. *Cognition, 124,* 72-78.

Moxley, J. H., & Charness, N. (2013). Meta-analysis of age and skill effects on recalling chess positions and selecting the best move. *Psychonomic Bulletin & Review, 20,* 1017-1022.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, *25*, 111-164.

Saariluoma, P., & Kalakoski, V. (1998). Apperception and imagery in blindfold chess. Memory, 6, 67-80.

Schultetus, S. & Charness, N. (1999). Recall vs position evaluation revisted: The importance of position-specifice memory in chess skill. *American Journal of Psychology, 112,* 555-569.

Simon, H. A. (1974). How big is a chunk? *Science, 183*, 482-488.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist, 19,* 394-403.

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

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

Waters, A. J., Gobet, F., & Leyden, G. (2002). Visuospatial abilities of chess players. *British Journal of Psychology, 93,* 557-665.

Witteman, C., & van Geenen, E. (2010). Cognitive process analysis. In A. Glöckner and C. Witteman (Eds.) *Foundations for tracking intuition: Challenges and methods* (pp 45-60). New York: Psychology Press.

BIOGRAPHICAL SKETCH

Jerad Moxley graduated from Murray State University with a B.A. in History and Psychology in 2005. He received his M.A. in experiment Psychology from Murray State University in 2008. His research is on the processes of expertise and individual differences related to performance. He has published in journals in the following fields: clinical psychology, sports psychology, and cognitive psychology.