

FLORIDA STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

REVIEWING EXPERT CHESS PERFORMANCE: A PRODUCTION-
BASED THEORY OF CHESS SKILL

By

ROY W. RORING III

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

Degree Awarded:
Summer Semester, 2008

The members of the Committee approve the Dissertation of Roy Roring defended on July 14, 2008.

Neil Charness
Professor Directing Dissertation

Bud Fennema
Outside Committee Member

K. Anders Ericsson
Committee Member

Colleen Kelley
Committee Member

Joyce Ehrlinger
Committee Member

Approved:

Janet Kistner, Chair, Psychology

Joseph Travis, Dean, College of Arts and Sciences

The Office of Graduate Studies has verified and approved the above named committee members.

In many ways this work culminates many years of education through which I have become the person I am today. Therefore, I would like to dedicate this work to my mother and father who always provided me the love, the encouragement, the opportunity, and the will to succeed and grow—the real purpose of any education.

ACKNOWLEDGEMENTS

My journey into graduate school was largely influenced by specific questions, and my arrival at FSU a result of seeking out the experts. I consider myself very lucky. Many graduate students do not find an advisor who not only exhibits wisdom and practicality, but also understands how to create the right environment for independent thought. I cannot easily express how often I have been grateful that Neil Charness is my mentor and how many times his guidance and his patience helped me during my tenure here. He has always directed me to new opportunities and his good nature has made graduate school a very rewarding experience. I have also been privileged to work closely with Anders Ericsson. There is no doubt that he has influenced my thinking tremendously. I will always remember our countless discussions and debates and how they provided the excellent opportunities for feedback and improvement in scientific thought. It was my outstanding good fortune to interact with such powerful thinkers.

I will also never forget how I have met some of the best friends of my life in the last several years. In particular I would like to acknowledge Edward Cokely, Mark Fox, Ainsley Mitchum, and Katy Nandagopal—I have learned so much from each of you and will always cherish your friendship. And finally I could never forget to mention my parents, Winn and Tibbie Roring, their unwavering love and support makes everything possible. After many years of education, I still can only hope to one day be as wise.

This work was supported in part by NIA grant 1 PO1 AG17211-06 and NIA grant 5R01 AG13969.

TABLE OF CONTENTS

List of Tables	vii
List of Figures	viii
Abstract	ix
1. INTRODUCTION.....	1
General Abilities.....	3
Chunking Theory.....	10
SEArch Evaluation and Knowledge.....	20
Long Term Working Memory.....	24
Template Theory.....	27
A New Theory of Chess Skill.....	31
2. STUDY 1: Predictions for Search and Practice Variables.....	44
Study 1: Method	48
Study 1: Results	50
Study 1: Discussion	55
3. STUDY 2: Predicting Search-Skill Correlations for Specific Problems	57
Study 2: Method	58
Study 2: Results	58
Study 2: Discussion	62
4. STUDY 3: Historic Increases in Chess Skill.....	63
Study 3: Method	64
Study 3: Results	66
Study 3: Discussion	70
5. GENERAL DISCUSSION.....	70
APPENDICES	72
A INFORMED CONSENT FORM.....	72
B IRB APPROVAL LETTER.....	73

REFERENCES	74
BIOGRAPHICAL SKETCH	85

LIST OF TABLES

Table 1: Example Search Tree	45
Table 2: Search Correlations and Descriptive Statistics	51
Table 3: Pattern Matrix for Factor Analysis of Search Parameters	52
Table 4: Pattern Matrix for Current Practice Activities	54

LIST OF FIGURES

Figure 1: Productions and Memory Representation in PRESTO	36
Figure 2: Problem 12.....	60
Figure 3: Declining Tactical Error Rates of World Champions	67
Figure 4: Serious Age Varies with Tactical Error Rate	68
Figure 5: Serious Age as a Function of Accumulated Solitary Study.....	69

ABSTRACT

Explaining expert chess players' dramatically superior skill represents an outstanding unsolved theoretical problem for cognitive psychology. This review extends and re-evaluates the current state of theories on chess skill, highlighting the strengths and weakness of various theories ranging from general abilities to chunking theory to more modern variants. After describing limitations with earlier approaches, a new theory is described based on Productions RELating STored Organizations (PRESTO) of chess pieces in memory. I discuss how this theoretical framework, extending and elaborating Ericsson and Kintsch's (1995) proposal for long-term working memory in chess, addresses established empirical findings, makes testable new predictions, and how it is related in several respects to the earlier theories. These predictions are examined in three studies, illustrating how the model can explain characteristics of chess problem solving and planning behavior as well as a historic increase at the highest level of skill.

INTRODUCTION

Near the beginning of the 20th century, psychological studies began to investigate the high levels of skill demonstrated by individuals in real-world domains (e.g., Bryan & Harter, 1899; Binet, 1894). These studies discovered evidence for superior performance levels beyond those observed in more traditional laboratory settings. Feats of expertise, such as playing multiple games of chess blindfolded or mentally computing products of very large numbers, are often difficult to explain using theories of learning and cognition derived from more simplistic tasks. One virtue of expertise research is that it pushes cognitive theory beyond its comfort zone, forcing theories developed from human performance on unpracticed laboratory tasks to accommodate and explain the extreme achievements of certain individuals. As this review will suggest, during the last four decades, early models of cognition struggled to handle the empirical findings from research on expert-performance. Research on expertise has subsequently led to improved methodologies for understanding human performance in a variety of domains ranging from ballet dancing to bridge play to wine tasting.

However, after early studies, such as Binet (1894), Bryan and Harter (1899) and Cleveland (1907), over half a century passed before more refined attempts to investigate high levels of skill evolved. In his pioneering research, de Groot (1965/1978) investigated the thought processes and memory performance of chess grandmasters and club-level chess players and described many qualities of the thinking and search characteristic of these high levels of achievement. His work was originally published in 1946, and in English in 1965, when the cognitive revolution was in full bloom. With psychology returning to the study of mental thought processes and to a computer metaphor of human cognition, many researchers, such as Herbert Simon, Allen Newell, and other prominent scientists, sought a model organism for cognitive psychology, much like the fruit fly (*Drosophila*) is a model organism for genetics.

Chess was already widely researched by computer scientists, at the time representing an outstanding problem in artificial intelligence: how to create a computer program that plays excellent chess. This must have added great appeal for researchers such as Simon, who later championed the notion that computer simulations of human thought represent complete cognitive theories, similar to systems of differential equations in physics. Inspired by de Groot's work, Simon argued that chess would serve cognitive psychology as this model organism (Simon & Chase, 1973), noting many important characteristics of chess. Finding the best move in a chess position is a highly complex, real-world human activity, and each chess position represents a well-defined problem environment, with a fixed number of identifiable moves that can be played at any given point—perfect for studying search processes and problem solving.

Chess has other highly desirable characteristics as well, including an objective metric of individual differences in skill¹.

Subsequent work by Chase and Simon (1973a, 1973b) first demonstrated the classic skill by typicality interaction for rapid recall of chess positions, where strong players show a memory advantage primarily when recalling (briefly-presented) chess positions taken from actual games but not necessarily for positions with pieces more or less randomly assigned to squares. Based on this and related empirical results, the authors argued for a theory of chess skill that incorporated key concepts from theories of memory, including Miller's (1956) notion of chunks stored in a fixed capacity short-term memory (cf. Atkinson & Shiffrin, 1968; Waugh & Norman, 1965; see Newell & Simon, 1972). Chase and Simon (1973b) described chunks as patterns of chess piece constellations in long-term memory with labels that could be stored in short-term memory during the recall of a position. Consistent with models of basic cognitive processes (e.g., Atkinson & Shiffrin, 1968), this theory allowed the "work" of the mind to occur in STM rather than LTM (STM is analogous to attention under these notions), given that long-term memories, being vast in number, cannot be directly retrieved but must be retrieved via a cue already stored in STM (Tulving, 1983).

The work of Simon, Newell, and Chase further offered a new paradigm for measuring and investigating expert-performance: if expertise is a function of stored chunks, then skill can be measured by comparing the memory performance of experts and novices on a recall task (as having more chunks should monotonically improve performance). Furthermore, as in chess, domain experts' superior memory should be constrained to typical situations and should show little, if any advantage in atypical situations. In fact, the skill by typicality interaction was later found in a large number of other domains, including the board game GO (Reitman, 1976), bridge (Charness, 1979; Engle & Bukstel, 1978), musicians (Sloboda, 1976), basketball players (Allard, Graham, & Paarsalu, 1980), computer programmers (McKeithen et al., 1981), and electronics (Egan & Schwartz, 1979). Simon argued that an understanding of skill in chess would provide insight into expert performance in a large number of other skills, arguing that such mechanisms, such as stored chunks and limited capacity memory stores, might explain performance in a wide range of different domains.

However, competing theories of chess skill have developed over the last three decades. In a recent review, Gobet (1998) compared different theories with data from chess memory tasks. This review compared the original chunking theory (Chase & Simon, 1973b), a theory based on knowledge and search (Holding, 1985), the theory of

¹ In the 1960s, Elo (1965, 1986) developed a system of numeric ratings for chess players based on the results of tournament games. These ratings increase when players win tournament games and decrease when they lose; moreover, the magnitude of the rating change depends on the strength of the opponent. The mean chess rating for tournament players is about 1,600 and the standard deviation is about 200. Players are often labeled experts on reaching 2,000 strength and grandmasters tend to have ratings above 2,500.

long-term working memory (Ericsson & Kintsch, 1995), and the template theory (Gobet & Simon, 1996a). Gobet (1998) concluded that “In addition to accounting for most of the data on chess memory, the template theory, as did the chunking theory, offers a comprehensive theory of expertise, including perception, memory, and problem solving” (p. 147). Given this relatively recent review of chess theories, it is important to clarify why a new theoretical review is needed.

Gobet’s (1998) review had several critical shortcomings. First, Gobet (1998) committed several inaccuracies in his description of opposing theories, culminating in a sequence of later exchanges in *British Journal of Psychology* (Gobet, 2000a; Ericsson & Kintsch, 2000; Gobet, 2000b). Ericsson and Kintsch (2000) highlight Gobet’s misunderstanding of their model when they state how they “agree with Gobet’s (1998, 2000a) extensive criticisms of this mechanism, but not with his claim that LTWM [long-term working memory] proposes generic retrieval structures as part of its mechanisms” (p. 573).

Second, some of the empirical findings described by Gobet (1998) are incorrect; one example is his statement that, regarding long stimulus exposure times, the “recall of random illegal games [positions] brings their [chess experts] recall of piece location close to that of weaker players” (p. 145), which is inconsistent with Gobet and Simon (2000), who found that stronger players’ recall advantage in random positions grows with increasing exposure time.

Third, as will be clear in this review, several relevant empirical findings have emerged since this review was originally published, as well as other relevant findings and even theoretical perspectives that were ignored. Finally, Gobet’s (1998) review extensively draws upon findings from performance on a chess memory task in evaluating different theories; however, evidence exists that this task is not representative of chess expertise, as will be discussed in this paper. In fact, Gobet has only recently begun an attempt to expand his theory of chess memory performance to chess skill more generally.

Overall, a new review is needed to fairly evaluate existing theories and frameworks against the appropriate and up-to-date empirical evidence for the full domain of skilled chess performance (see Gobet & Charness, 2006 for a more recent review of empirical chess phenomena). In this paper, I will review existing theories of chess skill, their key weaknesses, and ultimately I will propose a new theory of chess expert performance that explains the key empirical phenomena relevant to any theory and that suggests testable novel predictions.

Review of the theories and their primary sources of evidence

General Abilities. One of the original explanations of high level chess performance is that chess masters have greater general mental capacities or at least greater capacities regarding visualization, spatial skill, and/or memory. The effects noted by Binet (1894) of the surprising capacity to play blindfold chess as well as the large

individual differences in chess playing skill might be a product of chess players having greater mental powers in general. This attitude was reflected by British officers during World War II, who recruited chess masters to help them as code breakers and cryptographers. If an individual can play such masterful chess, the person might also be capable of other impressive mental feats, a view even supported by some modern researchers (e.g., Howard, 1999, 2005). However, studies attempting to find evidence that strong chess play requires high levels of general abilities have been consistently unsuccessful. A review of this research reveals very little support for these hypotheses.

The first study to investigate the mental powers of chess players asked very strong players and non-chess players to perform a variety of cognitive tasks (Djakow, Petrowski, & Rudik, 1927). Of all these tasks, these authors found that chess masters could only demonstrate superior performance on memory for chess positions. This is inconsistent with a general abilities hypothesis, which would have predicted superior performance for the chess masters. Another early source of evidence was a study by Baumgarten (1930) of the intellectual faculties of one of the most famous chess prodigies, Sammy Reshevsky (a.k.a., Samuel Rzeschewski), as an eight-year-old boy. Reshevsky later became the U.S. Chess Champion and even competed for the world championship, being one of the top chess players during the middle of the 20th century. As a boy, Reshevsky had toured the United States playing against adults in exhibitions, where he frequently defeated many of his (novice) opponents. However, Baumgarten found that young Reshevsky performed very poorly on tests of aptitude, such as verbal tests, performing worse than five-year-old boys from Berlin (!), though he performed above average on some memory tests. Given that Reshevsky had no education up to that point save learning the Talmud and some Hebrew, such findings could easily reflect this selective educational environment.

More recently, studies have found evidence that as a group, chess players tend to have above average IQ scores (Doll & Mayr, 1987; Frydman & Lynn, 1992). This might be interpreted in several ways; however, as we will discuss later, there is no correlation between chess skill and IQ within chess players as a group. The finding that chess players have above average IQ scores could support the hypothesis that playing competitive chess requires innately high levels of general mental capacities. However, this would not explain various counterexamples, namely chess masters who have IQ scores below 100. Reshevsky aside, the study by Doll and Mayr (1987) found that, as a group, strong German chess masters (ratings between 2,220 and 2,425) had a mean IQ score of 106.5 (significantly higher than 100, the population norm), but had a standard deviation of 7.5. Hence, their sample of master players must have included some individuals with lower than 100 IQ scores. The existence of such individuals shows that having a high IQ score is not a prerequisite for achieving high levels of skill. Hence, some other explanation is needed for why chess players tend to have above average IQ scores.

However, much evidence exists for at least one prerequisite to playing high level chess: the notion of deliberate practice. The concept of deliberate practice originated

as a distinction from domain experience and is defined by Ericsson and colleagues (Ericsson, Krampe, & Tesch-Römer, 1993) as particular activities or training that is designed to improve specific aspects of performance. In many domains of expertise, including chess, it is clear that many individuals do not improve even after years of experience. For instance, in chess many of the more experienced players in New York's world famous Marshall Chess Club have not improved in decades despite playing frequently throughout each year (e.g., Feldman & Katzier, 1998).

Similarly, casual golf and tennis players typically do not improve over time, and this finding is true in many other domains (see Ericsson & Lehmann, 1996). However, these individuals reaching plateaus in domain performance are probably not engaging in deliberate practice. In deliberate practice, aspiring performers must focus on the aspects of performance at which they are weakest, and engage in problem solving to master these areas. Because this involves frequent failures and high levels of concentration for extended periods of time, deliberate practice is not inherently fun or enjoyable. Most individuals reach a performance level they deem acceptable, and then cease engaging in deliberate practice, resulting in performance plateaus.

Ericsson et al. (1993) observed violin students at a prestigious music academy in Berlin, and investigated retrospective estimates of their daily deliberate practice before entering the academy. These authors found that the most accomplished music students showed a striking difference from the other groups: although all students engaged in roughly 50 to 60 hours a week engaged in music related activities, the best music students spent much more time (25 hours per week) engaged in solitary practice, compared to less accomplished music students who spent less time (10 hours per week). During solitary practice, these musicians reported high levels of concentration to master specific aspects of their performance, often based on teacher feedback in weekly meetings, thus meeting the definition of deliberate practice.

The finding that elite individuals spend more time in deliberate practice has been extended to many other domains, including chess. Charness and colleagues (Charness, Krampe, & Mayr, 1996; Charness et al., 2005) have demonstrated that the best chess players engage in consistently more solitary study than lower rated players and also own far more chess books, whereas frequency of tournament play was only a weak predictor and was generally not correlated after controlling for solitary practice (see also Pfau, 1983, who found no correlation; notably, Gobet & Campitelli, 2007 found evidence for "group practice" as a predictor, although this variable contained "practice with others" in addition to tournament games). Moreover, extreme motivation to reach very high levels of skill appears necessary to engage in deliberate practice and even predicts chess attrition (deBruin, Rikers, & Schmidt, 2007). Chess was the inspiration for the well-known 10-year-rule (Simon & Chase, 1973), as few chess players have ever reached the international level with less than 10 years of experience, and even some famous prodigies, such as former world champion Bobby Fischer, took a little under a decade to reach grandmaster level and only competed for the world

championship some two decades later (see Gobet & Campitelli, 2007 for a more recent analysis).

There is little doubt that deliberate practice is required for chess improvement, though the microstructure of practice activities has not been thoroughly investigated—typically players report studying collections of archived chess games in books to test their move selection prowess against the moves played by strong masters in these games, receiving feedback and new ideas. Subtle aspects from many studies have provided indirect evidence that strong chess players must be studying tremendous numbers of master-level games; often studies have small sections showing informal comments by masters after seeing positions chosen from obscure sources such as “there’s a vague recollection of a Fine-Flohr game” (de Groot, 1965, p. 324, italics added) or often the chess master will be able to recognize the position and even state which game it was chosen from (e.g., Chase & Simon, 1973b)! This is impressive considering the thousands (today even millions) of games that exist in published literature, that to be able to recall or recognize a specific position from master game (even if imperfectly) implies that thousands must have been studied.

In fact, de Groot (1965) offered some suggestions on the development of a chess player “based solely on the biographical data of a number of chessmasters, data extracted from the chess literature for the most part but also complemented by the author’s personal communications with [former world champion] Dr. Max Euwe and Hans Kmoch as well as his own information on and personal experiences with chessmasters” (p. 347). He claims that developing masters “devote an exorbitant amount of their time not only to playing but also to analyzing thoroughly every game of their own and often those of others; not to mention studying the theory of chess” (p. 348). Moreover, “by means of playing experiences and/or textbooks the player gets to know certain important general strategic and tactical rules; next, he learns to recognize and to handle exceptions to these rules—which in their turn grow into new, more refined rules; with new exceptions, etc. Finally, the player develops a ‘feeling’ for the cases in which these already highly specialized rules can be applied” (p. 351)². We will discuss de Groot’s notion of rules more thoroughly in the conclusion of this paper, but exactly how players study and encode these games has not been carefully examined.

Additionally, modern chess players also report spending many hours of daily study time on chess openings. Openings, though not well-defined, typically refer to the sequence of the first moves in a chess game that a player has played from memory (older textbooks tend to define it based on accomplishing basic strategic goals, such as

² de Groot (1965, p.351) continues: “Thus, for instance, the player learns how important it is to occupy the center of the board with Pawns; then he finds out that a too broad or too far advanced Pawn center may be weak since it may become an easy object of attack. Next the player discovers that advanced Pawns...can be both weak and strong...Finally, he develops an ever finer and more reliable feeling for the types of situations where the strengths and the types where the weaknesses of such a Pawn structure prevail.”

piece development and castling). Today, entire books are published on very specific openings, covering, for instance, all the known variations or plausible move sequences and ideas in a specific position occurring after 15 to 20 specific moves have already been presumably played. That some players have memorized this vast quantity of information speaks to their persistent and extended training efforts. Further potential evidence for the role of practice is the persistent correlation between starting age and playing strength (Doll & Mayr, 1987; Krogius, 1976), as younger players tend to have more time and opportunities for practice.

Another idea related to general abilities and chess is whether playing chess improves mental functioning and leads to higher IQ, which could explain why chess players have higher than average IQ scores. Hence, the below average IQ scores of some chess players would have been even lower before learning to play chess, according to this argument. However, evidence shows that chess masters do not have higher IQ scores than intermediate chess players, despite the strong evidence that they have engaged in thousands of more hours of practice. First, as mentioned earlier several studies have replicated a null finding, namely that chess ratings do not correlate with IQ scores (cf. Djakow, Petrowski, & Rudik, 1927; the BIS IQ test, Doll & Mayr, 1987; Intelligenz-Struktur test, Grabner, Neubauer, & Stern, 2006; Raven's Progressive Matrices, Unterrainer et al., 2006). Moreover, a study finding above average intelligence in chess-playing children, Frydman and Lynn (1992) found the average full scale IQ scores between 3 groups of increasingly higher Elo ratings did not differ significantly. Although one study found a raw score correlation with rating in very weak child players, it was not significant after controlling for grade level (as higher grade students would likely have higher raw scores; Horgan & Morgan, 1990).

A recent, but related study argued that chess skill correlated with IQ after controlling for practice variables in school children with lower ratings; however, the measures of chess skill were not rating, but rather a chess skill test that was partially based on knowledge of the rules of chess (in addition to a chess recall test: Bilalič, McCleod, & Gobet, 2007a). It is possible that higher IQ children in this relatively weak sample of players have greater knowledge of the many obscure rules of chess, independent of general chess skill; interestingly, this same study found no relationship with IQ in their "elite" subsample of children (mean rating = 1,603), and even found a trend ($p = .07$) for a negative relationship with IQ and their chess Elo rating!

The absence of the correlation with the same IQ test in adults (Unterrainer et al., 2006) and the finding that correlations between skill and IQ tend to diminish with longer professional experience has been shown in many domains (Hulin et al., 1990) including chess (Bilalič, 2006). The Frydman and Lynn (1992) study investigating French child players (mean Elo of 1,450) found a mean IQ score of 121 (on the French WISC), which is not too different from the Raven SPM IQ score of the highest rated player of all time, Garry Kasparov (Elo > 2,800), who scored 123 during the peak of his career (Der Spiegel, 1987). Notably, the Horgan and Morgan (1990) study administered the Raven IQ test to "seven adult masters and experts" where "six of the seven subjects (including

one 66-year-old) had scores at the 99th percentile” (p. 117, footnote). These players would, therefore, have Raven IQ scores higher than Garry Kasparov’s, despite having ratings more than 3 standard deviations lower.

To summarize, even though chess players as a group often have higher than average IQ scores:

- chess skill generally does not correlate with IQ
- one of the best chess players of all time has a lower IQ score than some much weaker players
- some famous chess prodigies and master level players have below-average IQ scores

Finally, there is even evidence that individuals with high IQ scores do not improve at faster rates. Horgan and Morgan (1990) found that the Raven IQ test did not correlate with improvement in rating after controlling for the grade of their child participants. Doll and Mayr (1987) found that the full scale BIS IQ test did not correlate with either a 1 year or a 2 year change in chess rating. There is even converging evidence from studies of older adults, who typically perform far worse on IQ tests than younger adults, do not show very large age decreases in performance. The age effects tend to be quite small and sometimes do not even show up in small samples (Elo, 1986; Charness, 1981b, 1981c; Roring & Charness, 2007).

Perhaps the best explanation of these findings is that individuals in higher socio-economic status (SES) and more academic environments are more likely to be exposed to chess and play in tournaments to obtain ratings. This also explains why some studies did not find superior IQ for chess players (relative to non-players) when the players and non-players were matched for education (Unterrainer et al., 2006). Moreover, this would explain why there are some very low IQ chess players and why improving chess skill does not improve IQ score. Tournaments, chess equipment, and membership into chess organizations all have financial costs as well as time costs, which is advantageous for higher SES individuals (with correspondingly higher IQ scores) with more financial resources to purchase learning resources, play in clubs and tournaments, and free time for practice. Moreover, individuals with higher IQ scores may be more likely to surround themselves with activities like chess, which are often considered intellectual.

Also, personality characteristics may lead specific types of individuals to play chess. Such an explanation is consistent with findings that individuals who take up chess have higher Big Five (Barbaranelli, Caprara, Rabasca, & Pastorelli, 2003) scores than those who do not, including greater openness, extraversion, and lesser levels of agreeableness (Bilalič, McCleod, & Gobet, 2007b; see also Charness, Tuffiash, & Jastrzembki, 2004 for evidence pointing more toward introversion than extroversion). Notably, openness is also typically correlated with IQ, meaning that it is possible that this personality characteristic is the cause of chess players’ often higher than average IQ scores.

Finally, although higher IQ is not a satisfactory explanation of superior chess performance, it may be that better chess players have superior general visuo-spatial abilities, better memory performance, or other higher abilities. Such an argument requires at least two findings, namely that as a group chess players are higher than the general population for the ability in question and that the ability correlates with chess skill. However, again the evidence does not support this.

Waters, Gobet, and Leyden (2002) found that chess players do no better than controls on visual memory tasks and cite a null result from earlier work (Lane & Robertson, 1979) that found no correlation between chess rating and spatial ability. This finding is consistent with Djakow, Petrowski, and Rudik's (1927) findings as well as null results between chess strength and other basic cognitive tasks. For example, Campitelli, Gobet, Williams, and Parker (2007) found no correlation between chess skill and simple reaction time, long-term learning of names and faces, and both verbal and spatial working memory span. Moreover, as we will discuss throughout this paper, chess players do not even exhibit much better memory performance for chess positions when the pieces are quasi-randomly assigned to squares. Overall, it seems unlikely that chess players are superior in any general cognitive ability and that deliberate practice best explains their superior performance and improvement over time³. The question to be answered by theories of skill is what exactly is changing during improvement and what mechanisms explain their superior performance.

Other Early Theories. One of the first steps in studying experts, intuitively, could be to find out what the experts think and how they believe they demonstrate their skilled performance. Though not rigorous as data, the suggestions of experts might stimulate plausible theories for how skill, particularly skill at chess, might actually work, given their experience in the domain. Not surprisingly, early investigators attempted exactly this (Binet, 1894; Cleveland, 1907). Without a doubt, these early investigations provided inspiration and motivated many modern ideas of later work on chess skill; however, the verbal descriptions of these authors, like the descriptions from players themselves, suffer from ambiguity and are not grounded in modern psychological paradigms.

Consider the following explanation for improvement in chess: "Progress in chess like progress in abstract thinking of any other kind consists in the formation of an increasing symbolism which permits the manipulation of larger and larger complexes" (Cleveland, 1907, p. 300). But what does "increasing symbolism" mean, what are these "complexes" and how do they produce superior move selections? In fact, today (and in the earlier work) a large number of the rival theorists (as well as empiricists) in chess skill are former tournament chess players—but many are still rival theorists. Chess

³ One mysterious finding is that numerical ability tends to correlate with chess skill (Doll & Mayr, 1987). Notably, some authors have claimed that IQ correlates with chess skill (Grabner, Stern, & Neubauer, 2007), although only numeric portions of the test actually accounted for skill variance.

playing authors have included Chris Chabris, Neil Charness, William Chase, Adrian de Groot, Dennis Holding, Fernand Gobet, Pertti Saariluoma, Herbert Simon, Han van der Maas, and Eric Wagenmakers among others (including myself).

Chunking Theory

To better understand the chunking theory, we will first review the early work that led up to this point, how it influenced generations of psychologists with a new approach to studying complex thought, and how its evidence inspired the concepts and data cited by Chase and Simon (1973b).

In his pioneering research, de Groot (1946, 1965) attempted to investigate the complex thought processes in elite chess players using the “think-aloud” method. In many ways, de Groot followed in the footsteps of many psychologists dissatisfied with associationist psychology’s trend toward investigating the simplest and most basic cognitive processes, and instead attempted to describe the structure of thought of individuals performing highly complex tasks. The think-aloud method described by de Groot asks participants to think out loud while they complete tasks so that their verbalizations can be treated directly as data.

This method was later scrutinized by Ericsson and Simon (1993) in a large scale review, who argued that with correct experimental procedure, no evidence exists contradicting the fundamental assumption of non-reactivity, namely that subjects’ cognitive processes do not change due to thinking out loud. They further argued for the treatment of verbal protocols as objective data, where the ideal approach involves roughly three steps.

First, a task analysis determines different process models for how the specific task can be completed. For example, in solving the math problem 25 times 36, a task analysis could reveal many strategies; one strategy might involve the traditional approach of multiplying 5 times 6, carrying the 3, and so on, whereas another strategy might transform 25 to 0.25 which is $\frac{1}{4}$, and times 36 is 9 making 900.

Given a list of different process models, a second step can be carried out, namely the development of a coding scheme. In the multiplication example, the experimenter might look for unique intermediate products. Hearing the numbers 15, 30, 18, 6, 180, and 72 would be expected from the traditional strategy but not from the second, whereas hearing 0.25, $\frac{1}{4}$, or 9 would be expected from the second strategy but not from the first.

In the final step, participants’ verbal reports are elicited using a simple procedure (see Ericsson & Simon, 1993 for more details) and their data is coded and analyzed. If one participant in the multiplication task verbalized, “Okay 25 times 36, let’s see, [pauses], carrying the 3, 180, okay it’s 900” the analyst would be able to rule out the second strategy thanks to identifying the 180, an intermediate product that would not have been generated using the fraction strategy. The protocols are used this way, to reject models, but not to prove them, given the possible incompleteness of any

given protocol. The strategies that *cannot be rejected* are potential candidate process models for the given task (see Ericsson & Simon, 1993 for other potential methods of analysis).

However, de Groot (1965) did not approach his think-aloud data quite this rigorously. His was a more descriptive approach, which aimed to describe the structure of thought processes rather than to identify candidate models (one might plausibly place his study as the first step in the above process, namely task analysis). Further, de Groot originally studied some of the most famous chess players at the time that had competed at an international chess tournament in 1938. In this early year, de Groot did not have access to tape recording equipment and sufficed to use pencil and paper to help record thoughts of the chess players and then to review them with the player directly after the task.

Despite his rough and ready experimental techniques, de Groot discovered several important characteristics regarding the structure of chess players thinking. He derived the first problem-behavior graphs (though the term was later coined and the procedure refined by Newell & Simon, 1965), which shows how the chess player considers sequences of ply⁴ during the task of choosing the best move from a single chess position. In his experiment, de Groot chose chess positions he had personally analyzed for years (from his own games), himself being a master-level player, and he was thus quite certain of the best solutions. Moreover, his positions, particularly position A, was solved by almost all of the grandmasters in his study, but by very few of the weaker players he recruited. These positions captured essential aspects of chess skill (viz., finding the best move in a chess position) and differentiated the best from the weaker players. However, one of de Groot's most striking findings about the differences in the thinking structure between master players and intermediates concerned his protocol data, namely that he found no significant differences in the structural characteristics of search!

Most of the protocols analyzed by de Groot contained repeated sequences of moves; for instance, "1.NxB, RxN; 2.take on Q5; for instance 2.NxN, PxN; wins a Pawn, but there may be compensation for Black on QN2. But better is 2...NxN; then 3.BxN, RxR is nearly forced, no, it is not, he can play 3...BxB as well" (de Groot, 1965, p. 441). From this sequence of moves (the actual protocol was much longer, this is only a snapshot), a problem behavior graph can be constructed. This graph is essentially a tree of variations (sequences of moves) considered by the player.

In problem behavior graphs, researchers often evaluate characteristics such as number of different base moves (a.k.a., episodes), maximum or average depth of search, number of evaluations, number of different moves, number of total moves, etc. (Newell & Simon, 1965; Wagner & Scurrah, 1971). From his data, de Groot (1965) discovered that most chess players, after an initial appraisal of a position, begin an

⁴ one "move" is defined in chess terms as a movement of both a white piece and black piece, and hence one move is two ply

“alternation of plans and with it the continual return to the original ideas [that] must not be construed as vacillation or wavering, however, since in the successive elaborations, *a progressive deepening and broadening of the investigation* is apparent” (p.106).

Hence, in order to search deeper into a position, chess players often restart their analysis of a given variation. The analysis of the general phases of chess problem solving marked an important step forward, and showed how masters often find stronger moves as a result of the search process; however, when de Groot analyzed his protocols for the number of base moves, the total number of moves, the moves per minute, and the depth of search, he found no significant differences between the strong and the intermediate players. This was surprising, especially as many people often assume some chess players are better than others because they think more moves ahead. However, de Groot found no evidence that grandmasters think farther ahead than intermediate players, nor did they appear to search more broadly or consider more total moves: “It is not generally possible to distinguish the protocol of a grandmaster from the protocol of an expert player solely on structural and/or formal grounds” (de Groot, 1965, p. 319).

To explain the individual differences in chess skill, de Groot noted that the grandmaster “immediately knows what it is all about, in which direction he must search; he immediately ‘sees’ the core of the problem in the position, whereas the [weaker] expert player finds it with difficulty—or misses it completely” (p. 320). According to de Groot, “if consequently the master can start thinking from a higher level, then this class [expertise] difference should come out clearly in the first few minutes, nay seconds, of the perceptual and thought process” (p. 321). Consistent with this notion, de Groot replicated an earlier finding that chess masters have superior memory for chess positions (Djakow, Petrowski, & Rudik, 1927), except using a very brief presentation time for the stimuli (mostly varying from 2 to 10 seconds). He found a sharp skill effect for recall of these positions, even at this very short presentation time.

The search structure from the protocols revealed important information relevant to a general theory of chess problem solving, but the absence of effects lead many later researchers, such as Chase and Simon, to argue that they did not play any role toward individual differences in chess skill.

Chunking Theory: Description. Inspired by de Groot’s (1965) work, Chase and Simon (1973a) tested a hypothesis that can be thought of as the converse of an empirical fact: More skilled chess players have superior chess memory, so might having superior chess memory result in greater chess skill? In other words, they hypothesized that perceptual recognition of patterns stored in long-term memory might underlie chess skill. Simon had already developed a model of human memory known as EPAM (Elementary Perceiver and Memorizer; Feigenbaum & Simon, 1962) that was similar to other models of memory. In fact, it shared many characteristics with the heavily influential modal model developed a few years later (Atkinson & Shiffrin, 1968),

including a short-term memory (STM) and a long-term memory (LTM), and was used to explain list memory effects.

The notion of STM used by Simon (and later authors) centered around empirical experiments a few years earlier (Brown, 1958; Peterson & Peterson, 1959), which discovered that after building proactive interference, individuals are unable to remember even a single newly-presented word (or consonant trigram) if distracted for a brief period of time (typically 20 to 30 seconds). Moreover, as argued by Miller (1956), STM is capacity constrained and can hold 7 plus or minus 2 “chunks” of information; for instance, 7 or so digits, meaningful groups of digits, letters, words, and so on, where chunks are stored in LTM (the mathematical constant pi might be a chunk in LTM for 314). The classic EPAM model set fixed parameters (derived from experimental results) for the amount of time to encode new information into LTM (between 5 and 10 seconds⁵; Simon, 1974), so according to EPAM, any information presented for 5 seconds or less would have to be stored in STM.

If chess skill derives from a collection of chunks stored in LTM, then skill at chess might consist of two primary uses of this store: selective search of a chess position via chunk recognition, and superior evaluation of positions also via chunk recognition. However, the notion of chunks in LTM formed a basis for the Chase and Simon model, and this was the assumption that the earliest empirical experiments began testing.

Chase and Simon (1973a) set out to test their hypothesis by presenting chess positions for 5 seconds, so that the chess masters would have to store the stimuli in STM (they would only be able to store 1 chunk in LTM in this interval), then try to validate that their superior performance (recollection of the placement of more chess pieces) would reflect 7 plus or minus 2 chunks of information. Reasoning that such meaningful chunks would have developed through the course of chess study, they developed two sets of stimuli, one with chess positions taken from actual chess games and another where the positions (originally taken from games) quasi-randomly re-assigned the pieces to squares. They tested three chess players (a beginner, an intermediate, and a master) and discovered the now well-known and often-replicated skill by typicality interaction for chess position recall (although the interaction was not yet statistically generalizable in this small N study). Here stronger chess players show their memory advantage only during the recall of typical (game) positions (not random positions), as the typical positions should be the only ones containing meaningful chunks.

The study went further and attempted to create a method for analyzing the content of these chunks. Chase and Simon observed and videotaped chess players’

⁵ This was derived from list experiments showing that total learning time was relatively constant despite variability in stimulus presentation times and number of trials per stimulus. Hence, the number of stimuli learned in a fixed amount of time yields around 5 to 10 seconds per item (8 nonsense trigram pairs in about a minute: Bugelski, 1962).

reconstruction of a position that was perceptually available (in plain view). They recorded the number of times that players looked back at the perceptually available position to reconstruct it on the board, how many pieces were placed between glances, and the between and within glance time intervals. They found that within-glance intervals were usually less than 2 seconds and that stronger players tended to take less time between glances. They argued that about 2 seconds is necessary to recognize a chunk, to label it, and then store it in STM⁶. Further, they investigated the number of relations in these chunks, and found that the groups of pieces within glances tended to have multiple relations (attack, defense, same color, same piece, proximity), unlike groups of pieces between glances.

The next step was to argue that pieces placed within 2 seconds during recall of a position similarly corresponded to chunks. Chase and Simon found that the size of the interpiece latency (time between two recalled pieces) was inversely proportional to the number of relations between the pieces and a strong correlation between the type and number of relations for groups of pieces within-glances and less than 2 seconds during memory recall, but that the relations of within-glance placements did not correlate with relations for latencies greater than 2 seconds during recall. Moreover, relations of pieces placed between glances correlated very highly with relations of pieces greater than 2 seconds. After demonstrating the high correspondence between the types of relations in the two conditions, they showed that the average chunk sizes (between 2 and 4 pieces) together with the actual recall performance indicated a capacity of about 7 chunks (most middlegame positions have between 21 and 30 pieces or so). They also found that stronger players tended to have larger chunks, by their definition, in the memory task and that most of these chunks were pawn chains, a castled king position, or clusters of pieces of the same color.

These findings led Chase and Simon (1973b) to develop a theory of chess skill. They argued that over years of practice, chess players acquire a large vocabulary of these chunks in LTM. Chunks are stored patterns of pieces-on-squares, meaning that a chunk is typically defined both by the pieces as well as their locations, which was supported empirically in later studies showing that transposing board quadrants decreases memory recall (e.g., Saariluoma, 1994). Notably, although Chase and Simon did initially allow that some chunks might exist independent of board locations, their simulation work assumed all chunks were stored with their specific board locations (Simon & Gilmartin, 1973). However, the chunks alone cannot produce chess moves. Chase and Simon argued that these chunks or patterns are associated in LTM to labels that in turn are associated to chess moves; in other words, that chunks are the inputs to productions which output moves⁷. The theory can easily explain the skill by typicality

⁶ technically defined, chunks, the raw groupings, are associated with labels in LTM, and these “labels” are stored in STM

⁷ A production is an IF –THEN statement, such as IF I see a king in check THEN I must move him out of check

interaction, given that chess masters have vast quantities of chunks for typical game positions (where the piece constellations will match the chunks in the master's memory) giving him a large short-term memory advantage. However, in random positions, where no chunks are apparent, the master will lose this advantage (though a slight advantage may still be evident from accidental chunks in random positions: Gobet & Simon, 1996b). According to the theory, when a to-be-recalled chess position is displayed, the player first scans over it and notices salient pieces (cf. Simon & Barenfeld, 1969), then specific groups of pieces around the salient pieces are recognized as chunks in LTM, with the corresponding labels of the chunks placed in limited-capacity STM. During recall, the player accesses STM directly, retrieving from the STM labels the LTM chunks, which are then recalled. Simon and Gilmartin (1973) developed a computer simulation of their model (MAPP; Memory-Aided Pattern Perceiver). From their simulation's performance on recalling chess positions, Simon and Gilmartin estimated that a master's "vocabulary" could easily extend to over 50,000 acquired chunks, similar to the number of words used in many languages. Overall, the original chunking theory is the most well-known theory of chess performance and is often cited in cognitive psychology textbooks (Charness, 1992). However, despite its apparent parsimonious and intuitive appearance, later research has considerably weakened the foundations of this theory.

Problems with this theory. The argument that chunks of pieces on squares are stored in LTM and associated with moves via productions is an appealingly simple and parsimonious theory of chess performance. Stronger players play better chess primarily because they have larger vocabularies of stored, increasingly larger chunks (pieces on squares), suggesting that de Groot's (1965) failure to find any differences in the search processes of differentially skilled players was due to better players' superiority deriving primarily from rapid recognition of the stored patterns. It is compatible with observations that elite chess players require years of experience, and it converges well with mainstream psychological theories of memory.

The strength of this theory was later enhanced when its prototypical empirical prediction, namely a skill by typicality interaction for domain recall, was observed in many other diverse domains, suggesting a broad, almost domain-general theory of expertise. Given its pervasive influence, the Chase and Simon chunking theory has received greater empirical scrutiny than most other theories of chess skill. However, under the microscope of empirical data, the theory proved at best overly-simplistic and perhaps largely wrong in many key respects.

Even in the original work, the chunking theory had recognized limitations in what it could explain. Chase and Simon (1973b) discuss several subsequent findings to their initial work on recall of positions, namely LTM for chess positions, LTM for chess games, immediate recall of move sequences, and a task known as the Knight's Tour. However, rather than explain each of these findings with the notion of chunks as constellations of pieces on squares stored in STM, Chase and Simon explicitly mention

additional mechanisms. They argue that “the differences in chess skill manifest themselves in the speed with which successive new chunks are retrieved from long-term memory: 3 or 4 seconds for the Master, 6 or 8 seconds for the Class A player, and about 12 seconds for the beginner” (p.256). Therefore, not only does the stronger player have more chunks in LTM, he or she can retrieve them faster. In another experiment, Chase and Simon (1973b) required the players to recall memorized games after delays, noting that experts “have hundreds and perhaps thousands of sequences of moves stored away in long-term memory. The top players have thousands of opening variations—some running over 40 plies deep—committed to memory. There are also hundreds, perhaps thousands, of traps and winning combinations of moves that every Master knows” (p. 261). Now the chess expert has more chunks, retrieves the chunks faster, and also makes use of stored move sequences. Though these stored move sequences might also be chunks of some alternative variety, the third experiment, which compared memory of random versus meaningful move sequences, found that skill effects in move-sequence memory for both coherent (meaningful) and random move sequences⁸.

Finally, Chase and Simon validated the existence of skill differences on the Knight’s Tour; however, it is highly unlikely that piece on square chunks could explain this skill effect, given that only 5 pieces are involved: 4 black pawns and a white knight⁹. Grandmasters can do this task faster than skilled intermediates (intermediate players often have many years of experience themselves), so simply being more familiar with the basic knight movements cannot explain this away. Overall, even from their first study, Chase and Simon’s (1973b) theory was less parsimonious than it appeared in later citations, including far more mechanisms than piece-on-square chunks and their associated productions.

Moreover, piece-on-square assumption of the Simon and Gilmarin (1973) simulation was inconsistent with other empirical results, such as a skill effect on the errors of commission. Holding (1985) points out that 75% of these errors of commission were translation errors, questioning whether pieces are tied to specific squares: “The translation errors might involve shifting an entire configuration, such as a pawn chain, or else misplacing a major piece like a R or B along its own line of control” (p.109). According to Holding, this would “suggest massive reductions in the earlier estimates of the number of patterns that must be remembered in order to support chess mastery” (p. 109). Also unexplained by chunking theory is an impressive phenomenon known even by Binet (1894), where chess masters can play entire games (sometimes multiple games) without looking at a chess board. Known as blindfold chess, this accomplishment often demonstrated by intermediate and strong players requires

⁸ The authors committed a methodological error by using the games of former world champion Lasker, which their master player would likely have already been familiar.

⁹ The task does not involve noticing constellations of pieces, but simply how fast a knight can be maneuvered.

access in memory to a tremendous amount of information and, by comparison, most novices are incapable of playing even a handful of moves blindfolded.

However, according to Chase and Simon, the master's STM is filled by a single position, so how can a master player make strong chess moves during blindfold chess unless consideration and evaluation of planned chess moves takes place in LTM? In fact, Chabris and Hearst (2003) found no significant differences between chess play during rapid blindfold chess and regular rapid chess, where the player has about 15 minutes to play an entire game. The master's STM would be completely filled at any given position, so the masters should be unable to calculate or plan, though this is not the case.

Other aspects from the original study allow for criticism, particularly with their proposed method of identifying chunks. In particular, if the successive placements of pieces within the 2 second boundary are chunks stored in LTM, the chess player should consistently select the same chunks across repeated trials. However, upon investigating this, where a "chunk was defined as intact on the second trial [of the memory task] if at least two thirds of its pieces were recalled together" (Chase and Simon, 1973b, p. 228). Strikingly by even this rather loose criterion, they found that "65% of the Master's chunks and 96% of the Class A player's chunks remained intact on the second trial" (p. 228). The authors do not report the proportion of the Class A player's reliable chunks defined by having all the same pieces (as would be expected by the theory); moreover, that only a little over half of the master's chunks are the same brings into question what these groups of pieces really represent.

Although Holding (1985) suspects that masters might make "creative new insights into a repeated position" (p. 108), this possibility would both weaken the Chase and Simon methodology, as chunk reliability cannot be verified, and also potentially add an additional component to a theory that is already growing in complexity. The method of identifying chunks also did not transfer well to another domain. In a study of memory for typical versus random positions from the board game GO (where the skill by typicality interaction was found), Reitman (1976) did not find correlations between relations in chunks identified from a copying task and in chunks identified by the 2 second boundary, unlike Chase and Simon (1973a). Reitman tried a large number of different boundary conditions, and failed to find the correlations in each case. If the chunking theory of chess is extendable to other domains, chunks in such domains should be identifiable by the same methods. Moreover, the chunks explicitly identified intuitively by experts often overlap, contrary to the assumptions of independent piece on square chunks, and later experiments showed that chess masters also segment positions into overlapping units (e.g., Chi, 1978; see also Gold & Opwis, 1992).

Later experiments even suggested that the representation of chess players' recall of a position is hierarchical in nature (Cooke et al., 1993), also violating the independence assumption of Chase and Simon. Finally, it is unclear how most of the chunks identified by Chase and Simon (1973a) would plausibly generate moves. For instance, often chunks fail to capture interesting relationships in the position as 75% of

the identified chunks were pawn chains, castled positions, and pieces of the same color—even Chase and Simon (1973b) “were a little surprised at the importance of these visual properties and, related to this, we were surprised that the players made so little use of the attack relation” (p. 232). In verbal protocols, chess players often report attacking relations and it is difficult to explain why they are so rare in chunks, and Holding and Reynolds (1982) even found that unmoved pieces (not verbalized as moves in protocols) were significantly more likely to be remembered than pieces actually moved! Moreover, whether “proximity” can really be considered a relationship between pieces is quite debatable (see McGregor & Howes, 2002). If only attack and defense are considered plausible piece relationships, much of the data linking copied chunks to recalled chunks is questionable. Overall, the 2-second boundary method lacks reliability, face validity, and generalizability for the identified chunks.

To support the validity of the 2-second boundary, Gobet and Simon (1998) replicated with a larger sample many of the analyses of Chase and Simon. However, these authors did not examine the most critical problem with the 2-second boundary, namely reliability of chunk placement. If chunks, as identified by the 2-second boundary, represent information pre-stored in LTM, chess players should consistently recall the same piece constellations from a given chess position, but this was not examined by Gobet and Simon.

Other potential issues arise as well. These involve an assumption that is difficult to test, namely that in the copying task “the assumption that one chunk is encoded per glance” (p. 230); for example, it is possible that multiple chunks could be encoded in a single glance and that this might vary with skill, particularly if stronger players have larger visual spans (Reingold, Charness, Pomplun, & Stampe, 2001) and require fewer fixations when choosing a move (Charness, Reingold, Pomplun, & Stampe, 2001). Also, master subjects may have studied the positions longer before placing pieces during the copy task (see p. 234) and one master did not even look back at the to-be-copied board even once (his data was omitted from the analyses).

One of the most compelling aspects of the original theory was its relationship to Miller’s (1956) magical number of 7 plus or minus 2 chunks as the capacity of STM. Defined by Chase and Simon’s 2 second boundary, the chess experts tended to remember about the right numbers of chunks; however, the master remembered about 7 to 8 chunks on average which was greater than the intermediate and beginner, who remembered between 4 and 6 chunks (although no statistical tests were applied). This finding is not consistent with other domains showing a relation between chunk size and memory span. Simon (1974) observed his own (verbal) memory span performance, finding that as the chunk size increased (viz., from a word to 2 words to a sentence), that his chunk span (immediate ordered recall) decreased. However, the chess master in Chase and Simon’s experiments had larger chunks and a larger memory span¹⁰.

¹⁰ Unexpectedly, Chase and Simon did not find the master having larger chunks during the copying task, an oddity given that both tasks were presumably involving chunks.

Furthermore, his span was greater than recent estimates for visual STM, of around 4 chunks (Cowan, 2001; see also Gobet & Clarkson, 2004).

But perhaps the most devastating findings for the assumption that chunks are stored in STM come from the different studies showing that this cannot be entirely true—according to many cognitive models, including Simon and Gilmarin (1973), information in STM is erased once it is replaced. Hence, an interfering task requiring STM between the encoding of a chess position and its subsequent recall should dramatically reduce recall performance if the chunks are stored in STM. However, a study by Charness (1976) found that various interfering tasks, including mental arithmetic and mentally rotating and copying abstract symbols, do not produce large drops in chess position recall for strong players. Moreover, a study by Frey and Adesman (1976) showed that players could actually recall two separate positions at a time, which should not be possible based on Chase and Simon's model where STM is filled up by just one (however, for one possible way to introduce an additional mechanism that could explain this latter finding, see Gobet and Simon, 1996a). These studies provided strong evidence that much of a chess position is stored in LTM, despite that the positions were presented at very fast rates (about 5 seconds). In fact, LTM variables, such as depth of processing, may also affect position recall (Lane & Robertson, 1979), further supporting this latter position.

In fact, it turns out even the skill by typicality interaction itself changes meaning at longer exposure times. One of the key arguments was that chess players perform no better than novices when chunks are absent, hence in random positions. Later replications of the typicality by skill interaction, however, revealed small skill effects for random positions (Gobet & Simon, 1996b), and although these authors attempt to argue that chunks are occasionally present even in these random positions, experiments during the last two decades re-affirmed earlier work (e.g., Lories, 1987; Djakow, Petrowski, & Rudik, 1927) that given 60 seconds instead of 5, large skill effects on recall are observed for random positions (Gobet & Simon, 2000). This deviates significantly from the original findings of the chunking theory and, again, forces the theory to add supplementary mechanisms to account for this. Moreover, the interaction vanishes both when a recognition test is used rather than recall (Goldin, 1979) and when different information intake tasks are examined, such as counting bishops and knights (Saariluoma, 1985; see also Saariluoma, 1990b), or when the position is presented auditorily (Saariluoma, 1989).

Also problematic are studies showing either that skill effects exist without memory effects or that memory effects exist without skill effects: a double dissociation. Charness (1981c) observed that a sample of older chess players with equivalent ratings to a younger sample had worse performance on a memory task. This was evidence for memory effects in the absence of skill effects. Moreover, Holding and Reynolds (1982) found that higher rated chess players made superior move selections

in pseudo-random positions, without showing a memory advantage¹¹. Hence, this possible double dissociation weakens the idea that any chunks possibly supporting the memory task also support better chess play.

Clearly, there are many problems with the original theory posed by Chase and Simon (1973b) and simulated by Simon and Gilmartin (1973). In summary:

- The assumption of STM use is not defensible, leaving the primary empirical findings unexplained
- The main empirical result, the skill by typicality interaction, does not hold at long exposure times where experts show large recall advantages
- The theory continually adds ad-hoc additional mechanisms to explain different chess phenomena, and primarily only explains recall of chess positions
- The memory recall task shows a double dissociation with chess skill
- Several of the most important chess phenomena are not explained by the theory, including blindfold chess
- The method of identifying chunks is highly questionable; it is not reliable or domain general, and it is unclear how any of the identified chunks could be associated with moves
- It is unclear how chunks are acquired during practice

Finally, one of the early results that motivated the chunking theory was de Groot's (1965) protocol analysis showing essentially no significant differences between elite chess masters and intermediate level players' search characteristics. This finding led de Groot as well as Chase and Simon to expect skill differences to emerge in the first few seconds of the task, inspiring the memory stimuli. However, the empirical relationship between search characteristics and chess skill may be quite different than assumed by these authors, and as will be discussed in the following sections, the current picture is not entirely clear. If chess players really do search more broadly and deeply or even more rapidly than weaker players, this would potentially provide evidence for different mediating mechanisms. Moreover, the search characteristics arise from the most representative task of chess skill, viz. move selection, which accurately captures in-vivo chess performance. The first rival theory for the chunking hypothesis focused heavily on search, as will be discussed in the next section.

Search Evaluation and Knowledge Theory

¹¹ Notably, Schultetus and Charness (1999) found evidence for a post move-selection recall advantage for experts. However, the original Chase and Simon studies derived from evidence more closely resembling the pre-move-selection recall data.

Description. The theory proposed by Holding (1985, 1992) argues that three components underlie chess skill, namely skill at searching, skill at evaluating, and chess knowledge. “In outline, [SEEK] states that the skilled player uses his knowledge to generate an efficient forward search with accurate evaluations. Differences in these three components of chess skill, rather than differences in specific memory for patterns, are what the model takes to characterize the dimensions that separate higher- and lower-rated players” (Holding, 1985, pp. 245-246).

Holding presents empirical evidence for skill effects on these three dimensions. Several studies show that stronger chess players are better at evaluating positions (Holding, 1979; Charness, 1981c). Moreover, I found evidence from de Groot’s (1965) protocols to indicate how chess players are superior at evaluating. For instance, during the search process of one position, both grandmasters and weaker players arrived at the same end point in their calculations. In this position (position A), although the weaker player sensed nothing promising in this position, for several of the grandmasters, this was already a strong enough position to decide upon the chosen move. This example provides an exceptionally clear case of when stronger players and weaker players both consider the correct solution and resulting position, but the weaker players poor evaluation skill is insufficient to realize the quality of the play. Both the weak player and the stronger players found this position, so by chunking theory both players must have had the chunks to generate the moves—but only the strong players evaluated it correctly. This claim is strengthened by a computer analysis of this position from de Groot (1965), showing that the depth to a clear (material) advantage is deeper than all of the players searched¹². Many of them did not see that it wins a material advantage, but rather that it only leads to a superior position and even said so in the protocols (i.e., that they couldn’t see a material advantage). Skill in position evaluation must be critical in this position.

Second, Holding points to research illustrating that stronger players search more deeply and more broadly than weaker players. Although de Groot (1965) found no significant differences between grandmasters and intermediate players, Holding (1992) points out that grandmasters did search significantly more efficiently, namely by searching more unique base moves per unit time (cf. Charness, 1981b). He argues that de Groot’s study had few subjects and thus very little power to detect differences, and he points out all the search variables showed trends for the stronger players in the appropriate direction. Moreover, he argues that even small differences in search parameters, like depth of search, would potentially indicate large differences in actual searching efficiency.

¹² This informal analysis was conducted by the author, a highly experienced player, using analyses of the computer program Fritz 8, which is roughly as strong as the world’s top players. In this position, black has several subtle defensive moves that require very deep searching to resolve and were not considered by any of the players.

In fact, later studies did find that stronger players had superior search! Charness (1981b) found a independent relationships (via regression) between chess skill and various search characteristics, including number of episodes, number of terminal nodes, number of total moves, number of repeated base moves, maximum depth of search, mean depth of search, and unique base moves per minute. Holding and Reynolds (1982) found that the initial evaluation of a position changed significantly only for stronger players. Finally, Reynolds (1982) argued that an analysis of de Groot's (1965) protocols reveals use of a homing heuristic (similar to those described by Newell & Simon, 1972) that is applied more frequently by stronger players. This heuristic describes a trend for a player to broaden the search after negative evaluations and shrink the search after positive evaluations.

Third, several studies have shown that chess players have greater general knowledge of the game (Holding & Pfau, 1985; Pfau & Murphy, 1988; van der Maas & Wagenmakers, 2005), and these are quite strong effects. It is well-known that strong chess players memorize long sequences of opening variations and many common endgame positions¹³. Moreover, Holding (1985) argues that knowledge is “actively invoked by players during the process of move choice” (p.246).

Holding (1985) explains the Chase and Simon (1973a) memory results by appealing to processes incidental to acquisition of chess skill: “Instead, it appears that chessplayers who actively process the given positions are able to integrate the general characteristics of these positions in a hierarchical, prototypical, or schematic format, not necessarily based on pairs of pieces, that constitutes an “understanding” of the positions. This is clearly sufficient to account for the skill differences between players and, with minor additional assumptions, might also account for a tendency to replace related pieces one after another” (Holding, 1985, p. 130).

Notably, Holding (1985) emphasizes that some aspects of chess skill might be captured by the chunking model, but that his model places far more emphasis on search processes: “In principle, there is no reason why a recognition-association theory should not coexist as a subsidiary component alongside a search-and-evaluation theory, although in practice there is a wide difference in emphasis between the two theories. One minimizes, while the other maximizes, the role of forward search. The two accounts of chess skill can be reconciled if one takes the view that the basic activity in chess play consists of exploring a search tree, but that the candidate moves for a selective search are supplied by a recognition-association mechanism” (Holding, 1985, p. 247).

¹³ According to Cleveland (1907), the advantage of memorizing opening variations is “evident. It enables one to place his pieces good positions relative to each other...to avoid disaster in the early stages of the game...it enables him to play with a minimum of effort during the early stages of the game” (p. 304).

Problems with this theory. One of the critical empirical predictions of SEEK theory is that chess strength should correlate with search characteristics, particularly depth of search. However, as noted above, the empirical findings are mixed. Although Charness (1981b) provided evidence for such a correlation, the interpretation of this finding is not immediately apparent. For instance, Charness (1989) reports a follow-up on one of his participants from the earlier study after the individual had improved from a weak to average tournament player to a strong master level player. However, Charness did not find any significant changes in the search characteristics for this individual who had improved in chess strength over four standard deviations! Also, a study by Wagner and Scurrah (1971) examined search characteristics of two expert level players, corresponding roughly to ratings of 2,000 strength (corrected in Charness, 1981b), which is probably about the equivalent of de Groot's stronger intermediate players. These players demonstrated the ability to search very deeply into a chess position, much deeper than many of de Groot's grandmasters.

Thus, searching deeply into a position is not a skill possessed only by grandmasters, and this supports the original interpretation of the de Groot (1965) analysis, which found no effect on search. These findings lead Charness (1981b) to suggest that depth of search plateaus after about expert level, as this study used players ranging from beginners to about 2,100 strength, so perhaps the skill by search correlation exists only in this range—the strongest players in de Groot's (1965) sample were outside this range and the weaker intermediates were perhaps close to a 2,000 rating. On the other hand, a recent study has found evidence that supports Holding's (1985) view. Campitelli and Gobet (2004) found evidence that depth of search (along with other search characteristics, such as breadth and speed) actually increases even up to grandmaster strength, arguing that mixed results from previous research derive from the stimuli used. In many chess position problems, deep searching is not required to find the best move, and many of the early studies may have used such positions. Notably, though, this study used a very small sample ($N = 4$), and the exact nature of the correlations between search characteristics and skill remain unclear.

However, Holding's theory does not link search, evaluation, and knowledge to basic models of cognitive psychology. For instance, do the search and evaluation processes occur in STM? Moreover, SEEK is not specified enough to link practice activities to the search, evaluation, and knowledge mechanisms. While studying openings might result in greater knowledge, it is unclear how evaluation skills are improved and search is extended from the solitary practice activity that distinguishes improving players (Charness et al., 2005); moreover, what prevents playing games for fun from improving search and evaluation? Additionally, how does SEEK explain chess players' ability to play excellent blindfold chess? Holding's (1985) theory cannot address these questions, lacking the required depth and detail. Without a connection to basic cognitive models, SEEK suffers from ambiguity and may be explainable with more elaborate systems based solely on chunks (e.g., Gobet, 1997 suggests how chunks might affect depth of search).

Long-Term Working Memory

Description. Given that a central problem with chunking theory was its reliance on STM as the place where pointers to chunks are stored and thus as the place where the cognitive “work” is carried out as an explanation of the chess recall data, later authors devised theories that presupposed that the cognitive “work” occurs in LTM. Ericsson and Kintsch (1995) coined the term Long-Term Working Memory (LTWM) to refer to this idea. In expert performance, they argued, performers develop a method of accessing LTM as working memory, instead of limited capacity STM, and they are able to do this by organizing relevant information into “retrieval structures.” Retrieval structures are objects stored in LTM that are basically a collection of cues that are associated with each other in a specific organization. Hence, only a single cue from context or stored in STM could access a retrieval structure in LTM, which then allows the performer to access relevant information in LTM via the cues.

Given that the primary source of empirical data for the chunking model was the chess recall task, and given that later evidence contradicted the chunking model’s assumption that chunks referenced in STM account for this task, the empirical bases for chunking theory was essentially undermined. Ericsson and Kintsch’s (1995) proposal for LTWM can additionally alleviate the problem of explaining the recall task by using a retrieval structure based on the 64 squares of the chess board (i.e. each square is a cue and the board is the organization of the cues). Thus, chess players’ encode piece locations and relationships between pieces (e.g., a knight on e5 attacks pawn on f7) and the piece locations in are stored in the 64-square retrieval structure in LTM (e.g., the retrieval structure has a cue for e5, which now is associated with a knight, and a cue for f7, which now is associated with a pawn, and, furthermore, the relationship between the two pieces is preserved—see Figure 4 from Ericsson and Kintsch, 1995). In other words, “A chess position is represented as an integrated hierarchical structure relating the different pieces to each other, and all pieces are associated with their corresponding locations” (p. 237). Now because the retrieval structure and, hence, information about the chess position is located in LTM, interference in STM will have little effect on its retrieval (Charness, 1976; Frey & Adelman, 1976). Moreover, the skill by typicality interaction is explained for both the 5 second presentation rate, and also for the 60 second presentation rate, where the nature of the interaction changes (Gobet & Simon, 2000), because at 5 seconds, recognizable relationships, which allow more rapid encoding into a retrieval structure, exist between pieces primarily in typical positions. Hence, typical positions can be more rapidly encoded into the retrieval structure, resulting in an expert advantage primarily for typical positions given only 5 second encoding times. However, with longer time to encode the position (e.g., 60 seconds), experts show an advantage even in random positions, having had more time to better associate information (less familiar piece relationships) with the set of retrieval cues. This is also consistent with Saariluoma’s

(1989) finding that skilled players show superior position recall in both game and random positions when positions are auditorily presented one piece at a time—piece locations can be encoded and later retrieved even when no meaningful configurations of chess pieces are present. Finally, once information is encoded into a retrieval structure, it can be rapidly accessed whenever required; for instance, Ericsson and Oliver (1984; see Ericsson & Staszewski, 1989) found that cueing a chess expert with any square of the chess board resulted in very rapid (about 2 seconds) and accurate (over 95%) retrieval of the information associated with that square.

However, Ericsson and colleagues have proposed that researchers move away from studying the memory recall task as evidence for skill theories. A primary reason for this is illustrated by a study originally conducted by Ericsson and Harris (1990), showing that a complete chess novice, with no knowledge of chess whatsoever, can reach master levels of chess position recall with only about 50 hours of practice specifically on the memory task. This finding has since been replicated by Saariluoma and Laine (2001) and by Gobet and Jackson (2002). Moreover, skilled chess players improve their recall performance fairly quickly after exposure to the recall task (e.g., Gobet & Simon, 1996b). Ericsson and colleagues argue that the central problem is that the recall task probably does not capture the actual expertise, but is instead an epiphenomenon—if experts spend thousands of hours training to improve their skill, why would a relatively brief period of practice result in large performance improvements? Ericsson and Smith (1991) suggest that researchers focus on tasks that are representative of the underlying skill. Such tasks would not show these rapid effects of brief practice (cf. Ericsson and Roring, 2007), and in chess the task that is most likely representative of chess skill is the move-selection task originally studied by de Groot. Notably, relative to move-selection and other chess tasks, chess memory is one of the weakest correlates of chess skill (Pfau, 1983; van der Maas & Wagenmakers, 2005).

That information is stored in LTM rather than STM is consistent with many different empirical findings, including the memory tasks (Cooke et al., 1993; Gobet & Simon, 2000; Lane & Robertson, 1979) as well as with several other findings. For example, it is consistent with the null correlation between IQ and chess skill, given that IQ is often highly correlated with short-term working memory (e.g., Ackerman, Beier, & Boyle, 2005), and STWM is essentially bypassed in LTWM theory. This is further consistent with the small effects of advanced age on skilled performance (Roring & Charness, 2007), as age tends to diminish STWM capacity with most working memory tasks. Moreover, preliminary evidence from neuroscience corroborates the general assumption of LTWM. Amidzic et al. (2001) asked chess players ranging from 1700 to grandmaster strength to play against a chess computer while recording gamma bursts, allowing them to identify activated neural regions. They found a strong relationship between playing strength and decreased use of the frontal and parietal lobes together with increased use of the temporal lobe, supporting the notion of that with increasing chess skill, players rely more on information in LTM (temporal lobe) than in short-term

or classical working memory (frontal lobe)¹⁴. Similar results were found by Campitelli, Gobet, Head, Buckely, and Parker (2007).

Importantly, the theory of Long-Term Working Memory is a general argument directed at skill in a broader sense and does not claim to be a specific theory of chess move selection. It is a cognitive theory that contrasts with models claiming that chunks stored in STM give rise to expert-performance by claiming that integrated retrieval structures in long-term memory contain the relevant information used by chess players during play. In short, chess expertise occurs primarily in LTM. However, the theory does not explain how information, acquired from deliberate practice, is refined over time and implemented in LTM to execute greater and greater skill in selecting chess moves.

Problems with the theory. Gobet (2000a) argues that LTWM predicts that chess masters should have superior memory for random positions than is observed. With five second presentation rates, strong chess players tend to have very little advantage over weaker players. However, as previously discussed this argument is unfounded. LTWM retrieval structures are not merely piece-to-square associations, but also encoded into LTM via relationships between pieces. Hence, a random chess position, having fewer obvious chess relationships would take more time to encode into LTM than a typical game position. Moreover, Gobet and Simon (2000) found that given increased time for encoding, the memory advantage of stronger chess players increases for random positions, exactly as would be expected by LTWM.

Gobet has further criticized LTWM in a broader sense, arguing that it amounts to verbal theorizing and is ambiguous in its testable predictions. According to Gobet and Simon (1998), “Ericsson and Kintsch’s, 1995, alternative proposal of a single hierarchical retrieval structure to store any type of chess position is not precisely enough specified to be tested against empirical data” (p. 228). In contrast, Gobet argues that computer simulations are required to fully specify the nature of cognitive processes, and that verbal theories, like LTWM, are typically too vague to make testable empirical predictions. Clearly, the proof is in the pudding in this regard, meaning that only future experiments testing aspects of LTWM will show whether Gobet is correct in this critique.

Overall, LTWM theory leaves open many possible ways of describing chess skill, its central theses being that skilled chess players store working memory information in LTM during play and that this is used and updated in a retrieval structure

¹⁴ This study cannot be easily interpreted in any of these contexts, though, given that each player played a unique game against the computer and consequently did not encounter the same stimuli, so we have no way to know whether the observed effects would exist if all chess players solved the same set of problems. Further, it is quite possible that stronger players use different strategies to play computers than weaker players, which would also damage the ecological validity of this task.

that is organized by the structure of the chess board and encoded via relations between pieces. Moreover, LTWM takes issue with the notion of using chunks to explain chess performance, primarily when chunks are defined as units of STM. However, LTWM does not fully reject that stored constellations of pieces in LTM contribute to individual differences in chess skill—rather than view these as specific patterns in memory that must be matched against perceptual features in a chess position, skilled chess players may use such stored information to quickly and efficiently discover piece relationships in a given chess position.

Template Theory

Gobet and Simon (1996a) attempted to address the primary experimental findings inconsistent with the former chunking model of Chase and Simon. Of particular interest to these authors was how to address the findings incompatible with the view that the recall data may be explained by chunk pointers in STM. Gobet and Simon provided further experimental evidence that chess players' recall cannot be based entirely on STM by showing how recall of multiple chess boards would require as many as 15 chunks in the STM of strong players in some cases, clearly outside the STM boundary of five to nine (Miller, 1956). Also, recall in their experiments exhibited curvilinearity suggesting a primacy and recency effect, indicative of LTM storage.

Without greater time to encode information into LTM (e.g., early models require about 8 seconds to encode new information into LTM, deriving from experiments with verbal materials, whereas board presentation times were generally about 5 seconds for many chess recall experiments), memory for briefly presented information must be based on STM according to traditional chunking models. However, if information could be encoded in LTM more rapidly, the excessive number of recalled pieces and the lack of interference in STM can be explained through recall from LTM. Gobet and Simon (1996a), therefore, add an extra component to the traditional chunking model, which they call "templates." Essentially, a template is similar to a traditional LTM chunk except that it (1) is typically larger than 5 pieces-on-squares and often derives from specific chess openings, (2) contains "slots" or squares which contain variable information (e.g., the square f3 might contain a white knight and f2 a white pawn, but f1 might be variable and contain a white rook or a white bishop; notably, other slots in templates can be more abstract in that they contain information related to plans or possible moves), and (3) these slots can be filled with information rapidly (e.g., 1 second), whereby this information is then part of the LTM structure.

Hence, information in a to-be-recalled chess position can be rapidly filled into a template, allowing better recall from whatever information remains in STM after an interfering task because each template (stored in LTM) requires only one pointer in STM. Hence, according to traditional chunking theory, an interfering task would wipe out most (not necessarily all) of STM, allowing very little recalled information or perhaps only allowing recall from contextual information. However, if more information

on locations of chess pieces can be rapidly stored in templates (i.e., in about one second), each of which require only one pointer in STM, then much more information can be recalled by accessing the templates. In general, templates are theoretically necessary to explain the STM issues with chess position recall and have the nice property of being analogous to other constructs in psychology such as frames (Minsky, 1977), schemas (Bartlett, 1932), or prototypes (Rosch, 1975).

Like its older counterpart, template theory places the most emphasis on fast perceptual, pattern-recognition processes in explaining chess skill. Evidence often cited as supporting the importance of pattern recognition typically takes form as skill differences in chess play remaining robust under severe time constraints. For instance, Burns (2004) found that 81% of the variance in normal chess ratings can be explained through speed chess ratings, where players must typically make about 1 move every 5-8 seconds. Similarly, Gobet and Simon (1996c) found that a top-level grandmaster still performed at a very high level despite having to make moves very quickly.

Simulations. De Groot and Gobet (1996) developed computer simulations based on template theory to explain quantitatively the chess recall data, naming this simulation CHREST (Chunks, Hierarchies, and REtrieval STructures). To illustrate how template theory is incorporated computationally and might derive specific empirical predictions, I will briefly review the mechanics behind the simulation.

CHREST perceives objects (chunks or pieces) in a chess position via a simulated eye. When possible, this eye uses the largest chunk met so far (the hypothesis) to fixate a new square. That is, it follows the branch below the chunk and fixates the square associated with this branch. When this is not possible, CHREST either fixates perceptually salient pieces (cf. Simon & Barenfeld, 1969), fixates squares followed by attack and defense relations, fixates new regions of the board, or makes random fixations. In this way, a group of pieces in a chess position can be presented to the simulation as objects.

The heart of CHREST is the modern variant of EPAM (Elementary Perceiver and Memorizer) network that was originally developed by Feigenbaum and Simon (1962) to explain empirical data from a host of verbal learning experiments. This network refers to the structure of LTM and how new information can be learned and stored in LTM. The structure of the EPAM network is a tree-like hierarchy of nodes serving as a discrimination network. In other words, at each node, one of several branches or “tests” (e.g., “Is there a white knight on f3?”) can be performed on an object to determine the next node in the sequence (i.e., set of tests represented by branches). At each node of the network is an image, and this is where LTM chunk information is stored. To access a specific chunk in LTM, an individual traverses a path from the root of the network to a specific image, and this path is defined by the specific elements contained in the object (i.e., chunk or single piece-on-square). Importantly, each node is numbered as it is created during learning (see below) for conflict resolution (i.e., when an object satisfies multiple branches/tests). Note that the path to a given chunk (stored in an image) need

not fully define the contents of that chunk. The path merely discriminates the specific chunk from all other chunks stored in LTM.

Furthermore, the EPAM network is grown by familiarization and discrimination. Any perceived object is presented to the net (via eye fixations), reaches a specific node in the network, and then is compared to the image of the node. If the image underrepresents the object, meaning that the elements in the image are a proper subset of the elements in the object, new features are added to that image (familiarization). If the information in the image and the object differ on some feature or some sub-element, a new distinct node is created at this endpoint (discrimination).

Moreover, template slots are created when (1) the number of nodes below a given node that share identical information is greater than three and (2) when this to-be-templated chunk has at least 5 elements (viz., pieces on squares). Finally, similarity links can be created between nodes and templates. During learning, CHREST first compares each chunk coming into STM with the largest chunk already in STM. If the two chunks are sufficiently similar, a similarity link is created between the chunks. This is used during later recognition, so that if two nodes have similarity links, CHREST selects the node with the most information value (i.e. with the most pieces-on-squares).

Problems with the theory. The template theory attempts to revive the influential chunking theories of Chase and Simon so as to correct the empirical difficulties of the early theory. However, in doing so the template theory inherits many of the same problems, a key example of which is the lack of parsimony. For instance, the original chunking theory did not attempt to explain a large number of chess phenomena with only the simple notion of perceptual chunks, but instead posited additional forms of stored information, such as stored sequences of moves, specific types of knowledge, knowledge of typical plans, rates of chess specific processing, etc.

Similarly, when Gobet and colleagues have begun applying the notion of template models to non-memory tasks, such as chess problem solving tasks, the model immediately grows in complexity. Gobet and Jansen (1994) developed a simulation model called CHUMP (CHUnks of Moves and Patterns) to try to simulate move-selection process with a computer program that plays very weak chess (partly because it involves only pattern recognition and no search processes). However, CHUMP grows two nets, one for patterns of pieces, like CHREST, and another for sequences of moves, and further these nets are connected by associative links, which are created when the chunk patterns contain the piece-on-square to be moved. According to de Groot and Gobet (1996), "In addition to nets for patterns of pieces and for (sequences of) moves, nets could be created for openings, plans, heuristics, tactical concepts, positional concepts, etc" (p. 245). In other words, individual differences in chess expertise cannot be explained with chunks and templates alone, but must also invoke stored move sequences, plans, tactical concepts, and many other constructs.

Notably, Gobet (1997) has developed a second simulation of move-selection called SEARCH, which begins to incorporate search processes (but does not search

more than one branch within an episode). Like CHUMP, this model adds additional cognitive complexity; for example, it incorporates a “mind’s eye” which is “a relational system that can be subjected to visuo-spatial mental operations and that stores perceptual structures, both from external inputs and from memory stores” (Gobet, 1997, p.293). The template theory alone appears incapable of parsimoniously explaining chess skill, supporting other authors’ criticisms of focusing empirical experiments on memory recall tasks (cf. Ericsson and Roring, 2007). The memory task must not be capturing the essence of chess skill if it can be explained without the many extra cognitive constructs that even Gobet and colleagues argue are involved. In this vein other phenomena, such as knight tour performance and blindfold chess play (e.g., Chabris and Hearst, 2003), do not appear readily explainable with only templates and chunks.

But perhaps the most serious issue with the template model is the consequence of its method of implicit acquisition of chunks and templates. Whenever the model discriminates an object with its network, learning processes automatically update the net. In fact, according to de Groot and Gobet (1996), in “the case of chess, it is likely that some learning occurs when players are trying to find a move in a given position” (p. 214). However, the problem is this: if individual differences in skill are based on quantity of chunks and templates and if chunks and templates are automatically acquired during encoding of chess positions and move-selection, then chess experience alone should predict chess skill to a great extent. But as previously discussed, chess experience is a very poor predictor of skill, and many chess players’ ratings do not change even after decades of consistent play in tournaments. Roring and Charness (2007) found that most active chess players’ ratings changed very little over the course of adulthood (typically less than 100 rating points) supporting related anecdotal accounts (e.g., Feldman and Katzir, 1998). The problem is that these active chess players should be acquiring new chunks and templates in every tournament game¹⁵, and should therefore be improving based on familiarization and discrimination of their networks, yet such improvement is rarely observed—improvement in chess requires deliberate practice.

It is unclear how the notion of chunks and templates can be reconciled with the concept of deliberate practice, which emphasizes the role of explicit learning processes in response to feedback (cf. Ericsson et al., 1993). Template models would need to incorporate the explicit nature of deliberate practice and its basis on feedback as a

¹⁵ One possible counterargument is that players are rarely encountering new chunk formations during most tournament games due to use of the same openings. However, given that most players lose games in every tournament (based on Swiss System tournament structure), if skill is based on chunks and templates, these losses are due to errors ultimately caused by lacking the chunks in LTM. Hence, some new potentially learnable chunks/templates would always exist and should still lead to skill improvement.

corrective mechanism to account for this real-world phenomenon in improvements of expert performance.

Finally, several further issues can be raised. Template theory is thus far unable to identify actual chess templates empirically, given the nature of the template slots. It continues to rely on the 2-second boundary for chunk identification, which was earlier argued to be problematic. For instance, Gobet and colleagues have attempted to examine “chunk size” based on this methodology (e.g., Gobet & Jackson, 2002), but have not demonstrated strong reliability statistics for these chunks. Without empirical observation of these chunks, empirical falsification of their characteristics and independent validation of their existence will be difficult, if not impossible. Moreover, the emphasis on fast pattern-recognition process over slower search processes has been challenged by van Harreveld, Wagenmakers, and van der Maas (2007). These authors argue using data from online chess play at different time controls and from time controls in FIDE world championship matches (1999 to present) that skill differences are diminished when chess players have less time to make their move selections (see also Lassiter, 2000).

A New Theory of Chess Skill

Earlier we described how de Groot viewed chess skill as more and more refined principles. As chess players improve, something must be stored in LTM to allow them to improve from deliberate practice. Also, LTWM provides a candidate mechanism for how information can be stored in LTM during actual chess play, avoiding the problems with STM discussed previously. Moreover, stronger chess players may somehow encode a chess position in such a way that they can cue the increasingly refined principles (e.g., LTM productions) to select potential moves through a process of planning and calculation. The refinement and acquisition of these principles would explain the need for deliberate practice during explicit skill acquisition, in contrast to template theory which allows implicit learning.

An important question is what are the major empirical findings that a theory of chess expert performance must ultimately explain? I submit that there are four primary phenomena, each of which is replicable and theoretically relevant. The first are the search characteristics during move-selection tasks, such as the correlation between depth of search and chess skill (e.g., Charness, 1981b). The move-selection task is a task representative of chess expert performance and, thus, captures the essence of chess skill. The second phenomenon that must be explained is blindfold chess performance, indisputable as an impressive and real phenomenon of memory and cognition (cf. Chabris & Hearst, 2003). Third, the theory must be compatible with the notion of deliberate practice, and how mere experience tends to predict chess improvement poorly (e.g., Ericsson et al., 1993; Doll & Mayr, 1987). Finally, the theory must be able to account for the memory recall interactions that formed the empirical backbone of the older chunking and template theories (e.g., Chase and Simon, 1973b;

Gobet & Simon, 2000). The success of any theory of chess skill will depend on its consistency with these four phenomena as well as its advancement of testable novel predictions. Naturally, a theory must ultimately be compatible with all empirical findings, but these robust and well-replicated data are central to any new theoretical approach.

A broad description of chess move-selection consistent with some earlier theories is that of apperception-restructuring cycles (Saariluoma, 1990a, 1992a, 1995). According to this general description, when skilled chess players attempt to find a good move, they first “perceive” a goal position, then attempt to close the problem space (i.e. find a sequence of moves that leads from the initial position to the goal position). The perception of such a problem space is termed “apperception.” When the problem space cannot be closed, they “restructure,” meaning they form a new apperception of the position and, thus, a new goal position, which they then attempt to reach via move sequences. Therefore, weaker players cannot find correct moves either because they do not find the correct problem space (i.e. do not apperceive the appropriate goal state) or because they are unable to correctly close the appropriate problem space (i.e. find a sequence of moves from the initial position to a correctly apperceived goal position).

Saariluoma’s work emphasizes the importance of problem spaces and this description is consistent with de Groot’s (1965) observation of “progressive deepening,” but it does not explain how stronger players achieve such apperceptive powers and precisely what processes lead to the apperception of appropriate problem spaces. But in general, the notions of pattern recognition, LTM encoding of chess relationships, production-based move generation, chess knowledge, positional evaluation, and problem space closure and goal states all have some form of support and intuitive appeal. Here I submit a new theory that incorporates many of these elements as parsimoniously as possible to explain the key empirical facts from chess research.

Proposal for a New Theory. In this section, I will outline a new framework for a theory of chess skill and its acquisition. I will begin by describing the general outline of the framework. Then I will describe how a chess player selects a strong move in a chess position based on the framework, how a chess player improves in skill, and how other findings are compatible with this view.

To begin, the framework assumes the basic cognitive architecture present in earlier theories, namely of a LTM that is accessed via cues in STM, where STM is limited in its capacity. While separate STM stores may exist for visual and for verbal/phonological information, I will be primarily concerned with visual STM in relation to chess play to store a cue to a LTWM retrieval structure; moreover, a central executive or high-level attentional component manipulates the contents of STM based on goals, initiates goal-related processes, and other high-level information; hence this architecture derives from Baddeley (1986). Tasks, such as dual-tasks or other complex tasks (e.g., random number generation), may tap the central executive, whereas basic memory span tasks, like digit span, primarily tap verbal STM. Importantly, empirical evidence suggests that the central executive is required during the selection of chess

moves from studies of dual-task interference, whereas verbal STM most likely plays only a minor role at best (e.g., articulatory suppression does not affect move selection) (Robbins et al., 1996).

At its heart, however, the model will further incorporate into LTM three sets of productions. These are POS-POS (piece-on-square) productions, allowing the creation of piece relationships and the generation of a LTWM retrieval structure, a set of structure-evaluation productions that evaluate the critical relationships in a position and determine which is most significant, and a set of goal-generation productions that can produce goal positions for players to attempt to reach through planning. As these productions are memory elements, they have the typical properties of memory elements, such as differing activation levels (e.g., increased from frequent use or decreased from disuse). This model will be fully elaborated below, and I will discuss why each component is necessary. The core of this theory is that piece relationships are the fundamental units of information used by chess players. As I will argue, these relationships are formed from knowledge of the rules of chess and of knowledge of general chess principles in the beginning of the skill acquisition process, but are later formed automatically using highly specific productions after extended deliberate practice.

I propose that upon seeing a chess position, several steps must take place. First, relevant relationships between pieces must be rapidly discovered (and stored in a LTM representation, namely the retrieval structure). Second, high level information must be activated to begin calculation and planning, with updates to the retrieval structure during this process. Finally, a move is selected. Hence, the model must describe (1) how a relationship is discovered in a position (and thus how multiple discovered relationships form a representation of the position), (2) how relationships are evaluated to determine goals for planning, (3) how moves are considered and eventually selected, and (4) how deliberate practice can improve this process.

How piece relationships are discovered and the construction of a representation. In my wording, I use the term “discovered relationships.” To clarify, relationships can be formed between any two or more pieces on the chess board, regardless of novelty; however, many such relationships would be irrelevant to a chess player. It need not be the task of the chess player to form as many relationships as possible, but instead to find a set of relationships that best facilitates move selection. In other words, the task is to find as many relevant relationships as possible. I argue that superior chess players not only construct such relationships more rapidly than do weaker players to more quickly reach a memory representation of the entire position, but they also are superior in finding more of the relevant relationships. As a consequence, superior chess players make fewer gross mistakes and blunders (e.g., Chabris & Hearst, 2003; Saariluoma, 1992a; van der Maas & Wagenmakers, 2005) in part due to not failing to find appropriate relationships (such relationships that can be then advantageously discovered by the opponent), and that they also construct a memory representation

more rapidly explains their superior memory recall of chess positions, which will be explained more thoroughly below. The following description will illustrate why this is the case.

Every chess player knows the basic rules of chess, and these rules are stored in LTM. From these rules alone, a chess player can form relationships in a position; for instance, a chess player knows the movement of a knight and can identify whether a knight is able to capture an opponent's pawn. Moreover, developing chess novices learn general positional principles (e.g., doubled pawns are weak) and thematic tactical motifs (e.g., pins and skewers) that help the player form relevant relationships. While solving a move-selection problem, a novice could identify a number of relationships in a position with sufficient time; however, not only does this process take substantial time, but storage of these relationships for use in planning is non-trivial for such a novice. The average chess middle-game position typically contains about 25 chess pieces, with a large number of possible attack-defense relationships. The beginner could adopt a number of strategies for maintaining these relationships in STWM, such as using verbal labels in verbal STM, storage in visual STM, or with effort an elaborate LTM for a set of relationships could be formed, but this set would be slowly accessed and subject to massive interference and forgetting, similar in many respects to a word list.

To adapt and improve, the novice must be able to (1) find these and other relationships more quickly and (2) store them more flexibly and accessibly. To accomplish the first task, the novice must somehow use information from previous chess experiences to predict what relationships will be discovered. Hence, information about possible relationships must be already stored in LTM, which I argue are stored as productions: IF piece P1 on square S1, piece P2 on square S2, ... AND other information THEN encode if piece PN is on square SN. The other information here could include previous positions or the specific opponent, but this factor typically plays a minor role and is primarily included for generality. The cognitive output of this production is twofold: it directs an eye movement toward SN and if the appropriate piece is found there, a relationship is formed, which I will denote for instance <P1_S1, P2_S2, ..., PN_SN>. I will refer to these as POS-POS productions, as they allow the player to encode relationships between pieces in a chess position based on stored LTM knowledge. Importantly, POS-POS productions can output a relationship between only two pieces or several different pieces. For instance, a production might be as complex as IF P1 on S1 AND P2 on S2 AND P3 on S3 THEN encode P4 on S4, resulting in a more complex relationship of <P1_S1, P2_S2, P3_S3, P4_S4>. This is possible when all POS-POS productions are learned initially via deliberate search for relationships in chess positions (by repeatedly applying rules, principles, etc.), which are unlimited in their POS scope.

This leaves us to explain the second task, how (and where) the player can store the new relationships and other information accessibly and flexibly. Consider the following example. When a beginner first discovers that a white pawn on e5 can take a black pawn on d6 (without help from POS-POS productions), the beginner must store

this relationship somewhere for the moment, but the fact that a white pawn on e5 can take a black pawn on d6 now has a long-term memory trace (though it might not be immediately accessible at first due to interference from other encoded relationships or due to lack of consolidation). Given repeated exposure to this relationship, the beginner's trace is strengthened and eventually becomes a POS-POS production, but importantly the beginner can connect the square e5 to the square d6 with this production—all specific pawn captures can potentially be associated with specific productions. Hence, now when a new position is analyzed, the production will fire as the beginner encodes a white pawn on e5. A resulting search of the square d6 revealing a black pawn will now create a <wPe5, bPd6> relationship stored in LTM. Suppose now that the beginner now encodes a <bPd6, bPc7> relationship—a more complex representation can be formed given the commonality of these squares, such as <wPe5, <bPd6>, bPc7>. Note that this more complex LTM representation associates each piece with a specific square of the chess board. With sufficient thinking time (and with the longer-term accumulation of POS-POS productions to speed up the process), the chess player can form a representation for an entire position that maps each piece to a specific square and preserves encoded piece relationships (the time to accomplish this based on the existence and nature of the POS-POS productions).

Ultimately the LTM representation uniting the squares of the chess board can be referenced with a single cue in STM as an organized retrieval structure to access locations of pieces (note that, particularly for weak novices, the retrieval structure may be less than 64 squares in some cases—this illustrates the generative nature of the retrieval structure as different from something fixed like Method of Loci journeys, as was confused by Gobet, 1998).

It is interesting to note that the POS-POS productions can explain the speed difference between an intermediate player and a master on the knight tour task: both players know how to legally move a knight, but the master has a more sophisticated knowledge of the chess board in terms of (knight) movements, such as knowing that Na1 is connected to b3 due to these productions. It is important to stress that these LTM productions (and therefore the ability to generate a retrieval structure rapidly) are deliberately learned as the player repeatedly searches for piece relationships during practice and play, but that if the player does not use the effort to search for new piece relationships, new productions will not be learned (notably, I suggest that some repetition may be necessary for an initially formed memory trace to an active production; however, whether a specialized neurological consolidation process is required goes beyond the scope of this initial theoretical formulation).

For instance, an intermediate player with a working knowledge of many productions for finding piece relationships may rely on the relatively effortless process of constructing representation from the pre-stored knowledge (i.e. the already-learned productions). Moreover, even when a player encodes a novel relationship during actual chess play, unless this same relationship is found again (in future games and positions), it will not become a POS-POS production (most games contain some

extremely specific elements that are unlikely to be repeated, particularly after the opening). In other words, during most game circumstances, an experienced chess player either activates existing POS-POS productions or encodes novel relationships that are too rare to be re-encoded enough in the future. Under these conditions, most experienced players will fail to learn new productions even after a large number of recreational or tournament games.

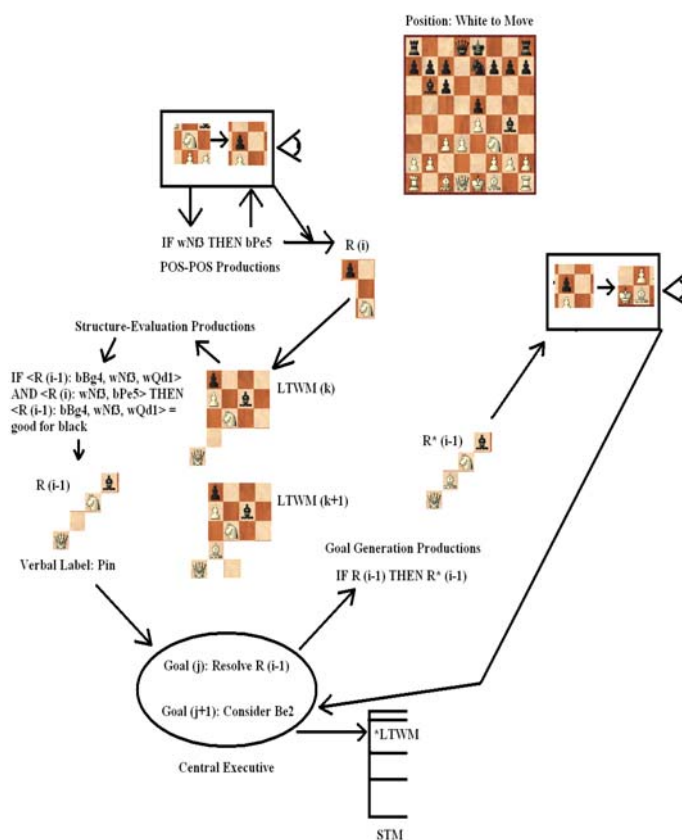


Figure 1: Productions and Memory Representation in PRESTO

Consider a concrete example of the POS-POS productions as illustrated by the position in Figure 1. In this position (white is to move), the player has previously encoded one relationships in LTWM (viz. $\langle bBg4, wNf3, wQd1 \rangle$), and now see a white knight on the f3 square. This POS serves as the input to a POS-POS production that outputs a black pawn on the e5 square and thus, with an eye movement inspecting e5, also outputs the $\langle wNf3, bPe5 \rangle$ relationship (note that without such a production in LTWM, the player would need to attempt to examine all possible movements of the knight

to find this relationship). This relationship then updates the LTWM representation as illustrated by subfigure LTWM (k).

It is important to recognize how this idea is similar and different from the chunks and templates of earlier authors. A chunk is a collection of pieces-on-squares in LTM (that is accessed in STM via a cue). The productions I describe relate pieces to each other in a similar way, particularly if other relationships are being used. The key differences are that (1) the productions (and hence the ability to generate a sophisticated retrieval structure for a given position) are explicitly, not implicitly, acquired, (2) chunks in LTM are “matched” to a complex relationship in a position in parallel, whereas productions “generate” complex relationships serially (but rapidly),¹⁶ (3) the chunks are actually pre-stored in LTM as complex unique patterns that are referenced in STM via one cue per chunk, whereas productions form an integrated memory representation of up to 64-squares that is accessed via a single STM cue. Notably, a template slot, where the player knows a set of possible pieces that could exist on a square, is not different in some respects from a set of productions, which could check for the existence of several possible pieces on a square. However, templates otherwise share the same theoretical differences as do chunks.

To summarize, as the player searches for new relationships in chess positions, those the player does not automatically find, productions are slowly formed over time. With a set of such productions, the skilled player encountering a new position can rapidly find many relationships and store them accessibly in a retrieval structure that maps pieces to squares and preserves piece relationships (a cue to the retrieval structure being stored in [visual] STM; see Ericsson & Kintsch, 1995). The more relevant productions possessed, (1) the more rapidly a player can generate an integrated representation and (2) the less likely the player will overlook important relationships contained in the position. This description is also consistent with the eye movements of chess players upon exposure to a position, given the tendency to direct eyes along piece relationships (Simon & Barenfeld, 1969; de Groot & Gobet, 1996).

Moreover, the description is consistent with other findings. Chess players may more rapidly count minor pieces even in quasi-random positions (Saariluoma, 1985;

¹⁶ Reingold, Charness, Schultetus, and Stampe (2001) argue that chess relations are encoded in parallel for expert players. However, the authors’ use of the term parallel is different than is used in this article: these authors state that “parallel encoding of chess relations by experts is hypothesized to occur following the initial extraction of feature information that is necessary to identify and localize pieces on the chess board” (p. 506, italics added). It is this initial extraction of feature information that POS-POS productions achieve, after which the relation is treated as a unit. Hence, there is no discrepancy with these empirical findings and the argument presented here. In fact, these authors’ findings offer support: chess players incurred a time cost in a check detection task when an additional piece was present (about 170 ms for experts), supporting the idea that an entire group of three pieces was not encoded as rapidly as a group of two pieces, hence not in parallel as the term is used here.

Saariluoma, 1992b) due to the use of complex POS to POS (piece-on-square) productions facilitating finding of the pieces. The famous skill by typicality by presentation time interaction (Gobet & Simon, 2000) is explained given that the quasi-random positions will not trigger productions as readily (e.g., such a position might have a king on f3, which is rare in actual play), so at very brief presentations (e.g., 5 seconds), the skill advantage in memory recall will occur primarily for typical positions, as in quasi-random positions with less typical piece placements the skilled player cannot as rapidly use the greater number of productions to his advantage. However, given more time, expert players can realize this advantage with their greater number of productions being serially activated, allowing more of their superior knowledge to play a larger role with more time. Also, the tendency for such memory recall scores to gradually increase with greater encoding time suggest a serial process of encoding rather than the parallel matching of chunks, which would produce a curve far more saw-toothed (see curves in Gobet & Simon, 2000).

Moreover, when quasi-random positions are presented auditorily, experts still show a recall advantage (Saariluoma, 1989), which fits well with the notion of productions helping predict and store relationships even in these quasi-random positions. Finally, these productions and their generation of a retrieval structure allow the player to play chess without sight of the board, as in blindfold chess. A better player, with a more sophisticated retrieval structure, can better maintain different piece relationships as they would exist on an actual board, reducing the difference between sighted play and blindfold play as playing strength increases (Chabris & Hearst, 2003).

Finally, that stronger players rely more on learned LTWM retrieval structures and less on storage of relationships in STM or in episodic list-like LTM is consistent with the finding that IQ and related ability tests decrease in predictive power as skill improves (Bilalič, McCleod, & Gobet, 2007a). Overall the description proposed here derives from earlier frameworks, such as LTWM (Ericsson & Kintsch, 1995), and it captures much of the intuitive appeal of chunks and templates (Gobet & Simon, 1996a).

Evaluation, planning, and search for the best move. As the player begins to establish a representation of a position capturing important piece relationships, the player must evaluate the relationships and generate possible moves addressing these evaluations. It should be noted that the representation of the position need not be complete before search processes begin, and often new relationships are updated in LTWM during search. In some cases, the discovery of a new relationship in the position during search may force the player to consider entirely different plans. Regardless, it is clear from players' verbal protocols and other findings that this search process allows the discovery of stronger moves and is an integral part of the problem solving process (de Groot, 1965; Chabris & Hearst, 2003).

Saariluoma (1992a, 1995) discusses how the player must work to reach a goal state from a given initial position or structure. He describes the processes of "closing" a problem space and when this is not possible "restructuring" the search to reach a new

goal state. However, I will argue that this is only one of two varieties of search processes. I will henceforth refer to this process as the closure search process, differentiating it from a process where players do not have a clear goal position in mind, but rather make several candidate moves and evaluate the resulting position. In the latter case, the player is not trying to reach a specific goal state, given that players will not always have a goal state, but instead is examining several potential moves and evaluating what is found. I will refer to the latter search process as evaluative search.

In order to search, candidate moves must be generated, but such moves are worthless without some method to determine the quality of positional structures arising from such moves. Beginning chess players lacking any knowledge of how to evaluate a chess position are likely to base decisions on attack-defense relations, particularly if they know the relative values of different chess pieces (e.g., a queen is worth more than a bishop). A beginner, after discovering a relation that a bishop can take a queen, may immediately play this move without further consideration or, given time and energy, may attempt to examine the consequences of this play. In this important latter case, an evaluative search process is initiated. The player will view the square on which the queen will move and examine further whether other pieces can attack this square. If, for instance, the player discovers that opponent's knight can capture the queen there, the evaluation of gaining a bishop but losing a queen (typically 3 pts. – 9 pts. = -6) is bad and the player rejects this line of play. However, if the player could better generate attack-defense relationships with this square, the appropriate evaluation could be more rapidly and accurately determined. This is particularly true if the search process extends several moves deep and entails multiple squares of interest. In this case, having a more sophisticated retrieval structure becomes critical to keep track of many different attack-defense relationships at once.

As the beginner improves, many types of structures are frequently encountered both in play and in instructional chess books. Pawn structures, relative locations of major pieces, locations of both kings, existence of certain pieces (e.g., queens) among other factors are important considerations, as are tactical structures, such as pins, potential double attacks, overworked pieces, and trapped pieces to name but a few. The player must not only identify the existence of these relationships, but must also evaluate them as favorable or unfavorable. The accuracy of the evaluation process requires extensive, refined knowledge specific to different types of positions that goes beyond the overly-generic nature of evaluations found in basic chess principles. For instance, in some positions, a specific pin might be very devastating for one player, whereas in others it might play only a minor role. Through both reading instructional chess books and through experimentation in chess play, an improving player slowly develops the skill at correctly evaluating these structures, not only in isolation but relative to other relationships in a given position.

This evaluative knowledge forms the higher level understanding of chess and is represented by higher level productions, which I denote structure-evaluation productions, of the form: IF relationships R1, R2, ...,Rn exist THEN assign relative

value to each relationship and return the most critical relationship R_i . Notably, these relationships correspond both to all relationships in a position, both those found deliberately and those triggered by POS-POS productions. To output the R_i relationship, each of the R_1 through R_N relationships receives an evaluation in the context of the other existing relationships, and R_i is the relationship with the greatest “good for me” or “bad for opponent” evaluation. The relationships may be associated with verbal labels in some cases, like “doubled pawns” or “pin on the knight,” which often emerge as verbalizations in think-aloud protocols. Such verbal labels exist due to the nature of the learning processes, such as the learning of general principles (e.g., Nimzovitch, 1930). For example, a beginner may evaluate doubled pawns as negative (a possible general principle)—this incurs an intermediate step, where a discovered relationship must be identified as “doubled pawns” before being evaluated as negative. However, with greater experience, specific relationships (more rapidly encoded using POS-POS productions) these general principles quickly lose utility, given that the evaluation of any piece relationship is highly specific to a particular position (e.g., some doubled pawn formations are often more good than bad—what matters for evaluation is not that they are doubled pawns, but what specific doubled pawn formation it is and what other relationships exist in a position). For the novice, these structure-evaluation productions will initially be relatively simple in nature, possibly focusing on evaluation of a single relationship or a handful of relationships (IF R_i THEN evaluate and return R_i). Once returned by the structure-evaluation production, the relationship R_i is transferred to the central executive as a goal: to resolve this relationship.

Once the goal of R_i resolution is initiated, a third set of productions, goal-generation productions, are activated. These productions take as input the R_i relationship and generate a goal position, denoted R^*_i . This allows the player to engage in closure search, to now augment the central executive goal state to look for pieces and examine moves that create R^*_i . If no such productions are fired, the player must instead engage in evaluative search unless a novel goal state has been created. However, if the player is now beginning closure search, the player’s central executive, by examining the viewed position for relevant pieces, considers moves by updating the LTWM structure and examining the memory representation (illustrated in Figure 1 by the LTWM (k+1) subfigure), triggering structure-evaluation productions to re-evaluate the position. Once the goal has been resolved, the search terminates and a move is selected.¹⁷ Notably, goals generated by these productions are similar in nature to the LTM constellations of pieces from chunking theories. The goal-generation productions are learned in LTM when players (1) create novel goals when solving positions and (2)

¹⁷ It should be noted that chess players, primarily beginners, occasionally use very high level goals different from those described here. These goals typically derive from learned general principles, such as “develop your pieces” and “trade pieces when up material.” As previously discussed, these principles are so often violated in more advanced play as they are supplanted by the specific relational productions.

learn specific maneuvers and move sequences from games in chess books and from game analyses. Hence, similar to the earlier argument for creating new POS-POS productions, an intermediate player will be unlikely to learn new goal-generation productions if the player avoids creating novel goals in positions or only creates novel goals that are unlikely to be repeated. As a result, playing recreational and tournament games is unlikely to improve skill once a working set of goal-generation productions exists.

Notably, the correct evaluation of relationships is absolutely critical! I argue that weaker players, aside from failing to discover relationships, often incorrectly evaluate and prioritize goals incorrectly. This leads many weaker players to consider searches of lines of play that stronger players would never even examine. Such blind alley searches are frequently observed in verbal protocols (de Groot, 1965). Moreover, weaker players may avoid certain lines of play because they incorrectly evaluate resulting structures as bad. Notably, this basic argument is generally consistent with Holding's (1985) emphasis on search and evaluation, and the productions can potentially explain the often contradictory quantitative findings from verbal protocols.

As previously noted, there is currently some evidence for a correlation between depth of search and chess skill, but in some positions, stronger players may search less deeply. Based on the proposed model, what is critical to determining depth of search is (1) what relationship has received top priority due to evaluations and (2) whether specific goal positions have been generated for closure search. Often differentially skilled players are not prioritizing the same relational structure and are actually engaging in different searches. I would expect a novice player to search more deeply than a grandmaster in some positions if the novice incorrectly evaluates a structure requiring extensive searching (e.g., with many piece exchanges) as the top priority. On average, however, stronger players will search more deeply for two reasons.

First, they will be more likely to engage in closure search, having access to a greater number of goal-generation productions. With a goal position to reach, greater depth must often be examined than with a search that merely evaluates different possible attempted moves. Second, as they can more rapidly identify relevant relationships, they will essentially have more relationships to evaluate in LTWM at any point in the search process. It is not an ability per se to search more deeply, but rather the identification of new problems (evaluated relationships) to solve as they consider lines of play (the LTWM retrieval structure keeping track of piece locations and relationships during this planning/calculation process). Hence, all three production types, POS-POS, structure-evaluation, and goal-generation productions can lead to greater depth of search on average for stronger chess players.

Moreover, I would expect stronger players to search with greater breadth on average, for this same latter reason, that they will have more potential problems to resolve, but again there will always be positions where stronger players search less broadly than weaker players due to differences in evaluations and in the relationships

encoded. Finally, stronger players will calculate more moves per unit time on average given they more rapidly encode piece relations using POS-POS productions and more often generate goals for closure search with goal-generation productions. Unfortunately, current search data are mixed and a large sample with a broad range of skill has not been examined.

Overall, structure-evaluation productions are grown and refined through reading chess books and experimenting in actual analysis, and this constitutes a critical element in improving chess skill. Moreover, accurate feedback is critical for refining these productions, and access to extremely powerful chess computers and to chess coaches is useful in accurately evaluating specific structures. An intermediate player unable to evaluate properly, often due to relying too much on very general principles (cf. Nimzovitch, 1930), rather than on far more specific principles relative to a type of position (cf. Watson, 1998), will go down many blind alleys. Together with a slower speed and less accurate encoding of piece relationships and less knowledge of potential goal state maneuvers, the player will reliably underperform a very strong player with much greater quantities of deliberate practice (cf. Ericsson & Smith, 1991).

Given the skilled players' enhanced speed and accuracy at discovering relevant relationships, superior ability to evaluate and trigger potential goal states, it is not surprising that speed chess captures most of the variance in chess skill (Burns, 2004; Gobet & Campitelli, 2007). Better players find relationships faster and keep track of them better, and given sufficient skill differences, a strong player can find relationships at a speed superior enough to play with a time disadvantage (Gobet & Simon, 1996c).

Novel predictions of the theory. This theory is ultimately based on Productions Relating STored Organizations (PRESTO). To briefly summarize the major claims of PRESTO: (1) chess beginners initially learn chess rules, basic general principles, and tactical themes and use these to form relationships with effort, (2) more advanced chess skill is partially represented by POS-POS productions that allow rapid, automatic identification of relationships and the generation of a retrieval structure in LTM, (3) more advanced chess skill is partially represented by structure-evaluation productions that allow precise assessments of relationships and later prioritization for goals, (4) more advanced chess skill is partially represented by goal-generation productions that can generate potential goal states to direct a player's closure search, (5) all productions are acquired during explicit learning processes. That the theory is almost entirely production-based (albeit three different types of productions) in its attempt to explain chess memory, problem solving, and other phenomena gives it greater parsimony and power than traditional chunking and template models, which frequently incorporated additional mechanisms when going beyond memory recall data. Moreover, the theory is tied to basic cognitive models (Baddeley, 1986), unlike older models, such as SEEK.

The theory makes several novel predictions. First, superior search data can address the predictions that stronger players, on average, search deeper, broader, and

more rapidly, and that this trend should extend throughout the distribution of chess skill (rather than plateauing around 2,000 strength).

Second, there are predictions regarding the practice activities relevant to skill improvement. Specifically, players must engage in activities where they search for new relationships in positions and consistently test specific types of structures for improved evaluation. Moreover, specific practice activities should correlate with specific search characteristics, given that the search characteristics are a function of learned productions rather than domain specific talent for searching ahead in chess. Chess activities may be classified as those where feedback on evaluation of structures is possible, such as coaching, analysis of databases, reading chess books, and playing powerful chess computers, and those where evaluation feedback is far more difficult and less likely, such as playing in tournament or recreational games, playing speed chess, and playing other games similar to but different from chess. The theory presented here suggests much stronger correlations from the former group (particularly after regression analyses reveal independent sources of contributed variance).

Furthermore, practice activities that allow for forming novel relationships in a chess position and for generation of novel goals should likewise correlate with search characteristics. Additionally, chess beginners are more likely to show some improvement from playing recreational and tournament games. This is provided that they engage in explicit processes to find relevant piece relationships and to create novel goal states, whereas more advanced players can rely on already-acquired productions. It should be noted, however, that merely playing chess alone (e.g., without even later analyzing one's games) is far from ideal, as these fledgling players will have trouble getting feedback on incorrect evaluations and will not easily be able to experiment with different structures to learn appropriate structure-evaluation productions.

Finally, feedback for improved evaluation is critical for improving chess skill, particularly with the structure-evaluation productions. Regarding this, chess history has long shown dynamic changes in the types of positions professional chess players reach in their games (e.g., Watson, 1998) and some in the chess community believe that chess skill is rising over historic time, particularly at the highest levels. Given that evaluations of positional structures are often learned from published chess games, books, and other material (number of books owned even predicts chess rating: Charness et al., 2005), the evaluations for many structures is likely becoming more and more accurate over time. If more modern players learn the more modern evaluations of such structures, they will reach superior levels of skill than players of previous generations by having superior quality structure-evaluation productions. Moreover, greater quantities of and access to such material would facilitate the acquisition of chess skill for the same reason. Access to superior material, environments, and trainers would facilitate accurate learning of structure-evaluation productions. Hence, finding a rise in skilled performance at the highest levels of chess skill would be consistent with PRESTO, in contrast to other notions that equate chess skill more with the quantity of

chunks in LTM¹⁸ or with notions that one's ultimate level of chess performance is limited by inherited genetic factors.

The specific productions used by a chess player can be observed in different ways. First, given that players often verbalize piece relationships when thinking out loud during move selection, potentially new POS-POS productions could be noted by players' verbalizations of these relationships. Subsequent exposure to positions also containing the relationship should show evidence for stored productions; for instance, players should more rapidly identify this relationship than a control group and be more likely to solve the problem if the relationship is relevant to the solution.

Moreover, eye tracking could be used to test POS-POS production based process models for solving a given position. A set of POS-POS productions would make specific predictions for eye movements of players and observed eye tracking data could rule out specific models similar to the verbal protocols methodology. Likewise, verbalizations in protocols could be used to rule out specific process models of sets of evaluation and goal-generation productions for a given position. Therefore, the model described in this paper allows researchers to understand the thoughts generated by differentially skilled players for any given position. Better players would be expected, on average, to possess most of the same POS-POS productions as weaker players, and greater quality and quantity of structure-evaluation and goal-generation productions.

Testing the Predictions. To test the predictions of PRESTO, three different empirical investigations will be conducted. First, using a large sample of players, tasks, surveys, and chess positions, statistical correlations and multiple regressions will assess how well the theoretical patterns of effects predicted by PRESTO match the empirical data. Second, an analysis of properties of actual chess problems will determine whether the PRESTO model can predict whether specific chess problems will require greater skill to search at the necessary levels. Finally, archival data sets derived from chess databases, including all games from the world chess championship matches, will be used to test predictions of the model for high level chess skill.

STUDY 1: PREDICTIONS FOR SEARCH AND PRACTICE VARIABLES

As discussed earlier, previous studies have found mixed results in relating search characteristics to chess skill. In this sense used here, "searching" a position means considering moves (ply) in a position, rather than searching for piece relationships. Each of these early studies was limited either by a low sample size or rating range, and other findings (e.g., Charness, 1989) made the assessment of skill improvement to increased search less clear. The model presented here claims that search in a chess position is driven by discovered critical relationships. Since different chess problems

¹⁸ Notably, chunking theorist might argue that quality of chunks changes over time.

have different sets of critical relationships, this suggests that many positions are likely to be idiosyncratic and may require more searching or less searching to reach the correct solution simply due to the unique features in the problem, leading to potentially different correlations between skill and search. And as most of the problems used in former studies were different (and rarely were very many different positions examined, typically less than five), this could potentially explain the mixed findings.

However, the PRESTO model does make several predictions for the correlations between chess skill and search when the search parameters are aggregated across a large number of move-selection problems. Ultimately this derives from five assumptions of the model: (1) discovered critical relationships drive search, (2) stronger players discover more relationships on average at any point in the search, (3) stronger players discover these relationships more rapidly than do weaker players, (4) stronger players have better access to relationships during search, and (5) stronger players tend to generate more goal relationships, which drive closure search.

Search Parameters and Skill. To examine how specific search characteristics correlate with skill, consider Table 1, which illustrates how the different parameters of interest are calculated from a sample search tree.

Table 1: Example Search Tree

Sample Protocol				
1	e4	e5	Nf3 f4	
2	e4	d5	Exd5	
3	e4	Nc6	O	b6
4	f4	e5		
5	f4	d5		

Search Parameters
 Unique Base Moves = 2
 Unique Other Moves = 11
 Episodes = 5
 Terminal Nodes = 6
 Branchiness = $6/5 = 1.2$
 Null Moves = 1
 Max Depth = 4
 Mean Depth = $17/6 = 2.8$

For convenience, in this table a hypothetical search tree has been generated from the initial position at the beginning of a chess game. In this example, it is less important to know exactly what moves are being played from the notation (each column of the chess

board is labeled left to right with letters, a to h, and bottom to top with numbers, 1 to 8; N stands for knight and omission of a piece implies a pawn is played), but rather consider each as a separate ply (it may be useful to note that in the first episode of searching, this player considered moving a pawn to the e4 square, then his opponent moving a pawn to the e5 square, then himself now moving his knight to f3. The player then reconsidered the knight move for a pawn to f4 move instead).

As shown in the table, a new episode is initiated whenever the player returns to the initial position to begin considering lines of play. In this search tree, there are five episodes. Each episode begins with a base move, and this search tree shows two unique base moves and eleven unique other moves. Each line of search results in a terminal node (e.g., at Nf3, exd5, b6, etc.), and the first episode in this search tree has two terminal nodes, leaving a total of six terminal nodes for the entire tree. Branchiness is calculated as the ratio of terminal nodes to episodes (I will describe the relevance below) and the search tree shown here has one null move (in episode three). A null move means that the player skipped this ply in the line of search; for instance, in episode three the player considered moving his pawn to e4, then his opponent moving a knight to c6, and then, regardless of what move he would play next, his opponent would play his pawn to b6. Finally, the line of search that is deepest yields the maximum depth of search, which is four ply from episode three, and the averaged depth across all lines of search in this tree is 2.8 ply.

Each of these particular search parameters is relevant to the PRESTO model. First, based on the assumption that stronger chess players discover more relationships on average to drive search, greater total moves searched, greater episodes, and greater unique base moves are expected for higher rated players. In other words, since higher rated players will tend to find more relationships, they will need to consider these relationships during search, resulting in a tendency for consideration of more moves in general, as well as more episodes and base moves. Note that while in some positions, discovery of a critical relationship may lead to less searching for better players, on average across many problem positions, stronger players will search more.

Second, better players will discover relationships faster on average due to having specific POS-POS productions and goal-generation productions. Stronger players find many relationships automatically, thanks to having specific productions available, whereas weaker players must find these relationships more deliberately. Hence, better players will search more moves per unit time than weaker players on average.

Third, because of this increased efficiency at discovering relationships and flexible storage in a LTWM retrieval structure, better players will be able to identify new relationships to drive search from their LTWM representation, rather than having to restart from the initial position. In other words, they will be less likely to lose access to information in LTWM; therefore, they will not need to recalculate the same variation from the beginning (i.e. to double check their representation's accuracy) and can instead consider other lines of search at deeper search depths.

This implies that the ratio of terminal nodes per episode will be higher in stronger players, whereas weaker players will be more likely to restart a new episode to continue investigations. Hence, branchiness should correlate with chess skill on average. Similarly, having more rapid discovery of relationships at any point in the search, and also more goal structures to achieve, stronger players will search more deeply than weaker players. Finally, more null moves are expected with greater playing strength as they are often (but not always) the result of closure search; however, this effect may be less pronounced given that null moves can arise for other reasons. As one example, a player may actually incur a null move by failing to identify a relationship that would have generated a (clear or obvious) move in its place.

Notably, all these correlations are expected to hold throughout the distribution of chess players, rather than reaching a plateau. Hence, these correlations will be examined for both the upper and lower halves of the sample.

Search Parameters and Practice. As these search parameters are correlated with chess skill, I hypothesize that they also should correlate with measures of deliberate practice, which is the primary source of improved chess performance in the PRESTO model. Since encoding of chess relationships without use of productions (which is required for a specific POS-POS production to ultimately arise) is an explicit process, practice activities that allow for careful examination of positions for novel relationships are more likely to lead to acquisition of these productions. Moreover, activities that promote feedback and allow for testing of appropriate evaluations of relationships in the context of other existing relationships are central for explicitly acquiring structure-evaluation productions. Finally, analysis and understanding of the types of goal structures that can result from other critical relationships allow the formation of goal-generation productions. The appropriate activities should correlate both with chess skill as well as with the search parameters described above.

Hence, I hypothesize that activities that rely primarily on automatic firing of pre-existing productions, rather than explicit encoding of novel productions, will not lead to much chess improvement. Time spent in activities such as speed chess (where a player must play all his moves in less than a brief period of time, typically about five minutes) or casual play in clubs or with friends should not correlate with skill or with search parameters (particularly for non-beginners, which comprise the sample described here). These activities also minimize opportunities for feedback. In fact, engagement in these activities might even detract from time that could be spent in superior practice activities, especially for individuals who participate in chess clubs and other activities for this sub-culture.

In contrast, activities designed to improve specific aspects of performance, such as studying the opening, middlegame, or endgame, should correlate with chess skill and search. In these activities, players can potentially accomplish all three study goals, namely learning novel POS-POS productions, learning to evaluate relationships, and learning goal-structures to achieve in specific positions. Each of these activities should

correlate with chess skill. Whether these activities correlate with search depends on the positions used in a particular study, given the high specificity of acquired productions.

For example, if no endgame positions (where only a handful of pieces remain on the board) were solved by players in an experiment, endgame study would be a weaker predictor than in an experiment where endgame positions are being examined. In fact, selection of moves in endgame positions has been far less commonly investigated in previous studies of search parameters than selection of moves in middlegame positions. Had endgame practice been available for examination, it is likely to be only a weak predictor at best of search in these middlegame positions. However, given that better players will have both more endgame and more practice in the opening and middlegame, some correlation would still be expected. In this particular case, a regression analysis examining opening, middlegame, and endgame practice as independent predictors of a search parameter should not reveal any independent contribution from endgame practice after controlling for other types of practice.

Notably, it is important to consider that most players studying openings do much more than memorize sequences of moves. Studying the opening implies studying the typical positions that arise from the first ten to twenty moves of a chess game, the goals for these positions, and what types of structures are good and bad. Often players will include study of chess databases for examples of how to play specific positions arising for the opening. Hence, the high specificity of this practice is ideal for acquiring productions and is actually a method for studying middlegames. By comparison, middlegame study typically refers to solving chess problems, such as studying either tactical middlegame problems that focus on a theme (e.g., pinned piece) or applying general principles to specific middlegame situations. Both of these types of study should predict skill and search.

Method

Participants

The participants comprising the dataset for this study were tournament rated chess players ($N = 180$) from four different countries: Canada (Toronto), Russia (Moscow), Germany (Berlin and Potsdam), and the United States (Atlanta, Orlando, and Tallahassee). Participants were recruited between 1997 and 1999 via newspaper ads, personal contacts, and announcements at chess clubs and tournaments. Participants were at least 18 years of age, the mean age of the sample was 45 years ($SD = 16$ years) indicating a wide diversity of ages, and a stratified sampling procedure was employed to ensure approximately equal numbers of intermediate (1600 to 1899), subexpert (1900 to 2199), and expert (2200+) level players within three adult age ranges: 18-39, 40-59, and 60+ years.

A subset of tournament rated chess players also completed the problem solving task ($N = 157$) described later. A subset of these participants agreed to a diary study (N

= 130) discussed below. The participants completing the diary study were not unique in rating (or solution accuracy on problems), as this variable did not correlate with whether the player filled out a diary, $p > .05$. Finally, another subset completed the full questionnaire again five years later ($N = 54$) for a follow-up to investigate the test-retest reliability of the questionnaire.

Notably, this data was not collected by the author, who instead obtained it from colleagues at Florida State University (cf. Charness et al., 2005).

Materials

Chess Move Selection. The chess problem solving task consisted of 15 chess middlegame positions, selected to vary in difficulty (1/3 easier, 1/3 medium, and 1/3 difficult). The quality of different moves in each position was rated by an international master, and we additionally scored moves as accurate if the move chosen matched the first choice for the powerful chess-playing computer program *Fritz 8.0* (Chessbase, Hamburg, Germany).

Verbal protocols were transcribed and a problem-behavior graph was extracted for each problem solving attempt (Newell & Simon, 1972; see Table 1 for an example of a problem-behavior graph). Problem-behavior graphs can be numbered according to episodes, each marking when the player revisits the starting position after some analysis.

Chess Activities Questionnaire. The questionnaire was a paper and pencil survey that included items related to (1) demographic variables (rating, age, serious age, etc.) and chess developmental milestones, (2) cumulative chess solitary study, (3) current chess activities, (4) chess activities in the first year of serious study, and (5) attitudes toward chess playing. In reporting cumulative chess activity, participants reported estimates of time on chess spent per week beginning with the year they learned to play chess up to the current year. Accumulated chess activity was calculated by multiplying the weekly estimates by 52 to get an approximate number of hours spent each year, summing this across all reported years to get total accumulated hours of activity. Surveys given to German or Russian sites were translated into the respective languages with back-translation to English to check translation accuracy prior to administration. Surveys took between 30 and 60 minutes to complete.

Chess Activities Diary. The diary asked participants to mark practice times of various activities after a day where they engaged in any chess-related activity. This included activities related solitary or group activities as well as to playing chess for fun at different speeds (e.g., blitz chess versus normal chess play), playing against a computer, working on openings, middlegames, and endgames, working with databases, solving chess problems, and tournament preparations. Some of these activities, such

as studying the opening, middlegame, and endgame were assessed in two ways, namely as studying in a group of other people or by oneself.

Procedure

After signing a consent form, participants completed the cognitive and chess problem solving tasks as well as developmental data questionnaire (see Charness et al., 2005). Participants were asked to think-aloud while they solved the chess problems according to the procedure recommended by Ericsson and Simon (1993). They were given a practice problem initially to familiarize themselves with the procedure and to allow them to adapt to prompts (when necessary) by the experimenter to keep talking. They were then given up to 5 minutes per problem and asked to announce their choice of the best move at that point.

After participants completed the cognitive and chess-related tasks, they received a cover letter and a questionnaire either in person or by mail. Participants completing the diary study were given up to 3 months to report the first 7 days they engaged in chess-related activities. Participants were instructed to mark activity estimates at the end of that day while the times were still fresh in mind.

Results

Search Correlations

Problem behavior graphs and their associated search parameters were extracted from verbal protocols (as described above; see also Newell & Simon, 1972). Similar to previous research with smaller samples (Charness, 1981a), rating predicted episodes (Cronbach's $\alpha = .94$), terminal nodes (Cronbach's $\alpha = .96$), unique other moves (Cronbach's $\alpha = .97$), mean depth of search (Cronbach's $\alpha = .89$), maximum depth of search (Cronbach's $\alpha = .91$), and unique base moves per second (Cronbach's $\alpha = .87$), and it did not predict unique base moves (Cronbach's $\alpha = .82$). Unlike previous research, the increased power showed how rating does predict null moves (Cronbach's $\alpha = .85$) and branchiness (Cronbach's $\alpha = .78$), but it did not predict total time taken per trial (Cronbach's $\alpha = .96$). These correlations and their statistical significance are shown in Table 2.

Table 2: Search Correlations and Descriptive Statistics. Correlations larger in absolute value than .157 and .205 were significant at $p < .05$ and $p < .01$ respectively. UBM stands for unique base moves, and UOM for unique other moves.

	1	2	3	4	5	6	7	8	9	10	11	12	13
Rating	1.00												
Total score	.53	1.00											
Accuracy	.48	.83	1.00										
UBM per sec	.17	.03	.00	1.00									
Mean time	-.13	-.02	.02	-.76	1.00								
UBM	.08	.05	.03	.44	-.04	1.00							
UOM	.46	.31	.30	.50	-.25	.38	1.00						
Episodes	.34	.29	.24	.20	.20	.71	.63	1.00					
Term. nodes	.35	.32	.27	.08	.27	.54	.67	.92	1.00				
Null moves	.28	.22	.15	.18	.08	.38	.66	.66	.71	1.00			
Max depth	.44	.32	.31	.18	.03	.14	.82	.47	.57	.58	1.00		
Mean depth	.36	.27	.27	.20	-.13	-.07	.72	.18	.29	.41	0.91	1.00	
Branchines	.19	.25	.23	-.14	.24	-.12	.43	.22	.51	.46	0.55	0.47	1.00
Mean	2021.34	922.71	0.41	0.03	192.24	3.1	14.93	7.24	9.47	2.3	6.5	3.35	1.34
SD	249.5	110.95	0.16	0.02	82.39	0.88	8.8	2.89	4.67	1.26	1.9	0.89	0.24

To determine whether search correlates throughout rating distribution, rather than plateauing around 2,000 strength as hypothesized by previous studies examining only players below this level (viz. Charness, 1981a), these same correlations were examined above the median rating of the sample (median = 2,001). Despite the reduced power, most of the correlations remain significant, including unique other moves, $r(79) = .30, p < .01$, terminal nodes, $r(79) = .29, p < .01$, mean depth, $r(79) = .29, p < .01$, maximum depth, $r(79) = .41, p < .001$, and branchiness, $r(79) = .23, p < .05$. Unique base moves per unit time, unique base moves, and null moves were not significant, although episodes trended in the right direction, $r(79) = .21, p = .06$.

Finally, a principle factor analysis revealed three primary factors among these search parameters (eigenvalue = 1; promax rotation). Inspection of the pattern matrix (see Table 3) suggests a factor based on breadth of search (loading unique base moves, episodes, and terminal nodes), a factor based on depth of search (loading on mean depth, max depth, branchiness, and unique other moves), and a factor based on speed of search (loading on unique base moves per second and mean time). These factors reinforce the interpretation of the search parameters reflecting deep, broad, or fast search.

Table 3: Pattern Matrix for Factor Analysis of Search Parameters

	Depth Factor	Breadth Factor	Speed Factor
Unique Base Moves Per Sec	0.00	0.22	0.92
Total Time	-0.04	0.22	-0.85
Unique Base Moves	-0.35	0.87	0.24
Unique Other Moves	0.67	0.33	0.32
Episodes	-0.03	1.01	-0.08
Terminal Nodes	0.23	0.83	-0.24
Null Moves	0.39	0.50	-0.05
Maximum Depth	0.94	0.04	0.01
Mean Depth	1.02	-0.29	0.13
Branchiness	0.64	0.00	-0.31

Practice Correlations

Practice Activities from Questionnaire. Only specific types of practice during the first year of serious practice (FYSP) correlated with chess rating, namely FYSP solitary study of chess positions, $r(174) = .15, p < .05$, FYSP chess instruction, $r(172) = .24, p < .01$, FYSP group study of chess positions, $r(169) = .17, p < .05$, and (negatively) FYSP rapid chess play (rapid chess typically means players' each have about 30 minutes to play all their moves), $r(170) = -.16, p < .05$. FYSP reading for fun, tournaments, blitz chess (where players' have about 5 minutes to play all their moves), casual games, playing with chess computers, and solving chess problems were not significant, $p > .05$. In general, casual chess activities and experience were unrelated to chess rating (except rapid chess play, which was negatively related), whereas instruction and study of chess positions were positively related. Notably, none of the FYSP study activities significantly correlated with search parameters.

Similarly, only specific types of current practice activities correlated with chess rating, namely solitary study of chess positions, $r(175) = .32, p < .001$, current group study of chess positions, $r(172) = .28, p < .001$, current frequency of reading for fun, $r(174) = .19, p < .05$, current tournament activity, $r(172) = .23, p < .01$, and negatively with current casual game play, $r(171) = -.16, p < .05$.

In contrast, current blitz play, current rapid play, current play with chess computers, current instruction, and current solving of chess problems were not significant, $p > .05$. Thus, higher rated players tend to study chess positions more, but they also play in more tournaments, play fewer casual games, and read for fun more often. They are more active in general, and engage in some types of activities that, when accumulated, may not contribute to superior performance.

Finally, a regression equation including all FYSP and current practice activities from the questionnaire explained 31.4% of the variance in chess rating, although adding total accumulated solitary study to the equation contributed unique additional variance, $R^2_{change} = .04$, $F(1, 130) = 7.012$, $p < .01$. Notably, none of the current practice activities from the questionnaire significantly correlated with search parameters.

Diary Analysis. Examination of the diary data revealed that rating correlates with studying the opening with others, $r(134) = .24$, $p < .01$, studying databases, $r(134) = .41$, $p < .001$, and correlates negatively with playing casual chess for fun, $r(134) = -.30$, $p < .001$. Rating did not correlate, but trended in the right direction for studying the openings alone, $r(134) = .15$, $p = .07$. Rating does not correlate with playing speed chess, playing rapid chess, playing a computer, playing in tournaments, or private instruction. Interestingly, it also does not correlate with studying the endgame or middlegame with others or alone. Overall, diary variables explained 29% of the variance in rating in a multiple regression equation, where studying the opening with others, studying databases, and playing casual chess for fun (negatively) remained independent contributors to explaining significant variance. Interestingly, studying the middlegame alone emerged as an independent positive predictor and studying the endgame with others emerged as an independent negative predictor, $p < .05$.

Of the practice variables that correlate with rating, namely studying the opening with others, studying databases alone, and the negatively correlated casual chess, all three predicted unique other moves (respectively), $r(119) = .27$, $p < .01$, $r(119) = .20$, $p < .05$, $r(119) = -.21$, $p < .05$. Studying the opening and studying databases predicted episodes (respectively), $r(119) = .25$, $p < .01$, $r(119) = .19$, $p < .05$, and casual chess trended in the right direction, $r(119) = -.18$, $p = .06$. Studying the opening and studying databases both predicted terminal nodes (respectively), $r(119) = .21$, $p < .05$, $r(119) = .26$, $p < .01$. Casual chess negatively predicted null moves, $r(119) = -.21$, $p < .05$, and studying the openings trended in the right direction, $r(119) = .16$, $p = .08$. Maximum depth was predicted by studying the opening, $r(119) = .19$, $p < .05$, and databases and casual chess trended in the right direction being (respectively), $r(119) = .17$, $p = .05$, $r(119) = -.16$, $p = .08$. None of the other correlations were significant, although studying openings trended in the right direction to predict mean depth, $r(119) = .16$, $p = .08$.

Overall, chess players spent 97 minutes per day on chess activities overall, ranging from 5 minutes per day to 383 minutes per day. In solitary activities, on average chess players spent 5 minutes per day studying the endgame (range: [0, 69]), 6 minutes studying the middlegame (range: [0, 75]), 10 minutes studying openings (range: [0, 83]), 2 minutes studying chess problems (range: [0, 46]), and 4 minutes studying databases (range: [0, 104]). By comparison, chess players spend, on average, 12 minutes per day playing blitz chess (range: [0, 134]), 7 minutes per day playing rapid chess (range: [0, 150]), and 16 minutes per day playing casual and tournament

games (range: [0, 184]). A post-hoc comparison of the mean times for the solitary activities to the times spent in play was significantly less, $t(129) = -4.67, p < .001$.

In general, most chess players spent little time in solitary study, and rarely spend more than 1 hour per day engaged in any specific solitary activity, but spend comparatively more time playing for recreation and in tournaments. This is consistent with observations that many adult chess players do not improve throughout many years of experience (e.g., Feldman & Katzir, 1998). Also of note is that these values highlight a likely overestimation of practice times in the questionnaire. The questionnaire values suggest about 20 hours per week of time spent in chess activities (about 2.9 hours per day based on a 7 day week), which is clearly greater than the 97 minutes per day suggested by the diary data, mirroring findings from other domains (e.g. Krampe & Ericsson, 1996).

Finally, I conducted a principle factor analysis (eigenvalue = 1; promax rotation) for the current practice activities across the diary and questionnaire data, shown in Table 4 below. This table illustrates a general consistency between the diary and questionnaire data, supporting the assumption of players' correctly interpreting the measures, despite some possible minor discrepancies. Interestingly, position study alone was identified as a separate factor from solitary study, which appeared more specific in the nature of the practice activities. Also of interest, rapid play and blitz play appeared to load on different factors, and casual play loaded differently than these activities, suggesting that many different casual activities are possible that are not highly correlated with each other.

Table 4: Pattern Matrix for Current Practice Activities

Measure	Variable	Solitary Study	Instruction	Group Study	Blitz	Casual Games	Tournaments	Problem Solving Others	Rapid Play	Position Study Alone
Diary	Blitz	-0.05	-0.03	0.00	0.76	-0.04	-0.04	0.14	-0.10	-0.06
Diary	Rapid	-0.13	0.00	0.01	-0.06	0.19	0.06	0.06	0.33	-0.12
Diary	Computer Play for Fun	-0.05	-0.03	0.53	-0.03	0.34	-0.09	-0.12	0.00	0.16
Diary	Human Play for Fun	0.01	0.07	-0.03	0.50	-0.16	-0.35	-0.09	0.32	-0.06
Diary	Tournament Play	-0.07	-0.07	0.09	-0.05	-0.02	0.37	-0.15	0.16	-0.17
Diary	Instruction	0.01	0.98	-0.06	-0.04	0.02	-0.10	0.08	0.02	0.00
Diary	Endgame with Others	0.05	-0.05	0.71	0.05	-0.05	0.04	0.16	-0.04	-0.10
Diary	Middlegame with Others	0.00	0.10	0.91	-0.04	-0.03	0.07	-0.11	-0.01	-0.08
Diary	Opening with Others	-0.11	0.51	0.31	0.00	-0.05	-0.05	0.01	-0.01	0.15
Diary	Prob Solving with Others	-0.04	0.15	0.02	0.05	0.14	0.05	0.82	-0.07	-0.06
Diary	Endgame Alone	0.92	-0.04	0.07	0.00	-0.08	0.03	0.07	0.02	-0.10
Diary	Middlegame Alone	1.00	0.01	0.01	0.03	0.14	0.00	-0.11	-0.09	-0.07
Diary	Opening Alone	0.65	0.15	-0.12	0.00	-0.05	0.11	-0.13	0.03	0.10
Diary	Problem Solving Alone	0.44	-0.12	0.01	-0.11	0.00	-0.15	0.38	0.03	0.18
Diary	Databases	-0.02	0.04	-0.08	-0.06	-0.03	0.14	0.30	0.09	0.10
Quest	Position Study Alone	-0.09	0.06	-0.19	-0.01	0.01	0.09	-0.03	-0.01	0.75
Quest	Reading for Fun	0.01	-0.04	-0.06	-0.03	0.12	0.44	0.03	-0.02	0.11
Quest	Tournament Play	0.12	0.12	0.04	-0.05	-0.05	0.50	-0.02	0.08	0.04
Quest	Blitz	0.03	-0.02	0.00	0.94	0.17	0.14	-0.10	-0.09	0.13
Quest	Rapid	0.02	0.01	-0.05	-0.07	0.10	0.17	-0.05	0.82	0.11
Quest	Casual Games	0.05	-0.03	0.00	0.03	0.71	-0.10	0.09	0.34	-0.09
Quest	Chess Computer Play	0.00	0.03	0.00	0.04	0.77	0.05	0.07	-0.03	0.00
Quest	Instruction	0.07	0.86	-0.01	0.04	0.01	0.11	0.08	-0.01	-0.05
Quest	Problem Solving	0.14	-0.07	0.20	0.05	-0.11	-0.04	0.08	0.09	0.61
Quest	Position Study with Others	-0.03	-0.07	0.07	0.15	-0.12	0.69	0.24	0.15	0.01

Discussion

The majority of the findings were consistent with predictions by PRESTO; however, some were deviant. I shall discuss these in two parts.

Consistent Findings

As predicted by PRESTO, many search parameters correlated with rating, including episodes, mean depth of search, maximum depth of search, unique base moves per second, and branchiness. Moreover, I found evidence that these search parameters continue to predict skill even above 2,000 strength, indicating that the search-skill correlation does not plateau as speculated by earlier authors.

Moreover, the types of practice activities during the first year of serious practice predicted current chess strength, including solitary and group study of chess positions, chess instruction, and even rapid chess, which negatively predicted current chess skill. In other words, players who tended to study chess positions and received chess instruction during their first serious year, while minimizing time spent playing rapid chess improved most. This is consistent with PRESTO in that activities maximizing opportunities for feedback (e.g. instruction) and encoding new relationships (e.g. studying positions) allows the formation of new productions, whereas time spent in activities in which time constraints prevent explicit encoding and feedback (rapid chess) even reduced the level of attained chess skill.

Current levels of different practice activities as measured from the questionnaire data predicted chess skill, suggesting that stronger players currently engage in more study of chess positions and less casual chess play. Notably, this is consistent with the FYSP data. Moreover, the diary study revealed similar trends in current practice regarding casual play (also negative) and further revealed how studying the opening and studying chess databases were predictive of chess skill. Also consistent with PRESTO, time constrained activities such as blitz chess and rapid chess did not predict chess rating.

Finally, diary data on activities correlating with skill, namely studying the opening, databases, and reduced casual chess predicted episodes and depth. In general, evidence was found supporting that stronger chess players search more and this is related to engaging in practice activities allowing the explicit encoding of novel productions.

Inconsistent Findings

Notably, some predictions were not supported. Unique base moves was not correlated with skill, though this may be a result of relatively lower reliability of unique base moves

compared to other search statistics (except branchiness). This finding may also be explained by idiosyncratic features of the particular positions used in this study, and specific positions chosen a priori to require greater numbers of unique base moves for accuracy should reveal correlations.

Another surprising result from the PRESTO perspective is that solving chess problems did not predict rating in either current or FYSP practice activities. However, some types of “problems,” such as mate in N problems, rely on artificial positions unrelated to those that appear in normal chess games (the factor analysis suggests that problem solving categories loaded on multiple factors). If players interpreted the category this way, it would not be expected to correlate based on PRESTO as these artificial positions do not contain many relationships meaningful to normal chess play.

A further inconsistent result is evidence from the questionnaire that stronger players currently engage in more reading for fun than do weaker players. In contrast to current levels of casual play, which was even negatively correlated with skill most likely due to taking time away from superior training activities, reading for fun predicted superior chess rating. This suggests that there may be some small advantage for this activity, and post-hoc explanations are available; for example, it may be that reading for fun leads players to learn more about the practice activities and training regimen of stronger players, such as advice on how a developing player should study the games of masters which is often found in chess books. If true, then more reading for fun could improve the player’s metacognitive understanding of how to improve at chess, though notably this effect is small. Hence, while not predicted in advance, this effect can be accommodated by the PRESTO model. It should be noted that reading for fun did not predict rating independently in a regression analysis of current practice activities.

Perhaps the most critical inconsistency is the absence of correlations between questionnaire data and search parameters. However, comparison with the supportive diary data could clarify this finding. As mentioned previously, studying openings and studying databases were the primary positive predictors of search parameters, whereas studying the middlegames and the endgames were not correlated in the diary data (although studying the middlegames did predict rating in a multiple regression analysis and evidence pointed toward a negative relation with studying endgames). It may be that the best and strongest predictors of search are reflected by activities related to studying the opening and databases, but neither of these activities was included in the FYSP or current practice activities data (and these variables loaded on different factors in Table 4). Hence, there was general consistency between the questionnaire and diary data considering that somewhat different activities were examined. As previously argued, studying databases and studying the opening is related to search under PRESTO. Nonetheless, it is particularly surprising that studying positions was uncorrelated, but this may be merely an issue of the effect being relatively small and difficult to detect. Notably, there were no significant correlations contradicting PRESTO’s argument that superior search results from appropriate practice activities, and the diary data lent supportive data.

Discussion

The data are overall consistent with predictions outlined by the assumptions of PRESTO, and deviations can be considered minor. Moreover, from these findings, a first approximation of an ideal training regimen could be outlined for aspiring players. In the first year of serious practice, a player should study chess positions and receive chess instruction, while avoiding wasting too much time in casual or rapid chess activities. The player should continue to study chess positions as skill improves, and should concentrate on understanding chess positions arising from openings (and using databases) and probably should play in tournaments to allow study of one's games for purposes of feedback. Interestingly, it may be less critical for these more advanced players to study the middlegame and endgame and chess instruction may be less useful as well. This could be due to the greater number of specific situations that an advanced player must master, often far too many than can be helpfully analyzed with instructional help; moreover, many books on the middlegame and endgame focus on general principles more than specific situations, which are more relevant to advanced players. In contrast, a beginner can profit greatly from instructional assistance, given that fewer productions need to be acquired in order to demonstrate clear improvement. In general, the author's impression is that a player starts by mastering general, widely applicable knowledge (e.g. chess principles, very common situations, learned from studying positions and from instruction) to mastering increasingly specific knowledge (e.g. from specific opening variations and ideas from rarer positions found mainly in databases), and this is fully consistent with the fundamental ideas of the PRESTO model.

STUDY 2: PREDICTING SEARCH-SKILL CORRELATIONS FOR SPECIFIC PROBLEMS

As a model of chess problem solving, PRESTO should be able to make specific predictions at the level of specific chess positions. In my review I argued that idiosyncratic features of specific problems are a key reason why earlier studies found mixed results on the search-skill relationships. Given that PRESTO presumes greater search to be a function of having more relevant relationships between pieces in a chess position, and further that acquired productions allow stronger players to more easily encode and store such relationships, there are most likely methods for predicting whether a problem will show a search-skill correlation in advance, namely by identifying the number of relevant relationships. Thus, as skill increases fewer and fewer of such piece relationships will be overlooked, resulting in greater overall search and superior move selections. However, that these relationships must be "relevant" is also important—most interesting are cases where search improves the quality of the selected move.

In this study, I will examine the specific chess problems from study 1 and the magnitude of skill-search correlations for each problem. I will then attempt to predict which problems require superior skill for appropriate searching using characteristics of the problem relating to quantity of relationships. Specifically, I will determine the number of attacking pairs in each position as a first-order approximation to the number of relevant relationships. The PRESTO model predicts that higher rated players are stronger because of acquired productions that help identify such relationships in positions, and that search is carried out to resolve those relations evaluated as relevant. Hence, the greater the number of relationships in a given problem, the more likely a player will fail to encode some (and then search based on them) due to lacking the appropriate productions. Finally, I will examine in detail the individual problem that best predicts chess skill.

Method

Participants

The data on the move-selection task with think-aloud reports from study 1 will be used for analyses here.

Procedure

For each of the 15 problems, I will calculate (1) the search-skill correlation (primary dependent variable), (2) the search-accuracy correlation, and (3) the rating-accuracy correlation. Each correlation will be Fisher transformed to ensure normality. Search parameters under consideration will be episodes, maximum depth, and branchiness as these are theoretically relevant and were found to correlate with chess rating in study 1. Accuracy refers to whether players chose the best move as indicated by the computer program *Fritz 8.0*. Alternatively, each player's moves were scored by an international master, serving as a secondary measure of problem solving performance, though this was highly correlated with accuracy, $r(156) = .83, p < .001$. Finally, I will calculate the total number of pairs of pieces in each position where either a white piece attacks a black piece or vice versa. This will serve as an approximation for the number of relevant relationships in each position.

Results

In support of the PRESTO model, the Fisher transformed skill - maximum depth correlations for each problem was predicted by the number of attacking pairs, $r(13) = .55, p < .05$. However, episodes and branchiness were not significantly correlated, $p > .05$. To explain this, I examined which problems showed evidence that the search parameters in question were relevant for problem solving. In other words, I examined

on which problems a given search parameter (e.g. max depth) could predict problem accuracy.

For maximum depth, accuracy could be significantly predicted on problems 2, 3, 5, 8, 9, and 12, $p < .05$. However, for episodes only problem 10 showed a significant episode-accuracy correlation. Hence, searching more episodes only mattered for 1 of these 15 problems, so this probably accounts for why greater relations did not predict a skill – episodes correlation since there was virtually no variability among this set of correlations (i.e. they were all close to zero).

In contrast, branchiness does show significant correlations with accuracy for several different problems, namely for problems 2, 8, 9, 11, and 15, $p < .05$. However, branchiness may be more reflective of memory failures (i.e. needing to restart search in a new episode due to uncertainty of piece locations in LTWM) than discovering relationships to search, consistent with the original explanation given previously.

Notably, when I examined the skill –search correlations after selecting only individuals who chose the correct solution to the problem, none of the Fisher transformed sets of correlations (for the three search parameters of interest) were significantly correlated with number of attacking pairs, $p > .05$, though the reduced N for each search-skill correlation probably increased the error in estimating these correlational parameters (a similar result is found when restricting to inaccurate solutions). However, in cognitive science the practice of examining only accurate trials is typically conducted to ensure that each participant engages the same mental process model when completing a task—given that a participant carrying out the correct mental process will arrive at the correct solution, incorrect trials are discarded, particularly when (reaction) times for the mental process of interest are being examined. In solving these chess move-selection problems, the differences in search reflect differences in process, and these individual differences in search are of interest in this study, rather than the times for specific mental processes. Importantly, the chance of guessing the correct solution is very low in chess move-selection problems, which typically contain around 40 legal moves. Notably, none of the correlations from Study 1 are changed when only considering accurate trials.

To statistically examine the effect of idiosyncratic problems, I conducted a repeated measures ANOVA analysis of accuracy, using Elo rating as a covariate. This revealed a significant main effect of problem, $F(14, 2184) = 6.09$, $p < .001$, $\eta_p^2 = .04$, a significant main effect of rating, $F(1, 156) = 45.71$, $p < .001$, $\eta_p^2 = .23$, and a significant problem by rating interaction, $F(14, 2184) = 5.09$, $p < .001$, $\eta_p^2 = .03$. This suggests that the correlation between rating and accuracy changed across different problems, meaning that not all problems lead to equally strong rating correlations. To highlight the idiosyncratic nature of search, I then conducted similar repeated measures ANOVA for relevant search parameters. For max depth, I found a main effect of problem $F(14, 2002) = 2.76$, $p < .001$, $\eta_p^2 = .02$, of rating, $F(1, 143) = 31.71$, $\eta_p^2 = .18$, and a significant problem by rating interaction, $F(14, 2002) = 3.38$, $p < .001$, $\eta_p^2 = .02$. For episodes, I found a main effect of problem, $F(14, 2002) = 2.98$, $p < .001$, $\eta_p^2 = .02$, of

rating, $F(1, 143) = 19.84, p < .001, \eta_p^2 = .12$, and a significant problem by rating interaction, $F(14, 2002) = 3.95, p < .001, \eta_p^2 = .03$. For branchiness, I found a main effect of problem, $F(14, 2002) = 1.79, p < .05, \eta_p^2 = .01$, of rating, $F(1, 143) = 5.48, p < .05, \eta_p^2 = .04$, and a significant problem by rating interaction, $F(14, 2002) = 2.30, p < .01, \eta_p^2 = .02$. Hence, each of these search parameters' correlation with rating depends on the specific problem under investigation, which supports the idea that the advantage in search for stronger players depends on the particular chess position in question.

Also of importance for understanding how specific aspects of problems affect search is the distinction between problems with a clear best move (based on the analysis of the *Fritz* computer program; I will call these “tactical” problems) and problems where the best move is not as obvious (I will call these “positional” problems). To examine whether the various search parameters are different between these two types, I will use the 1.50 evaluation criterion of *Fritz* (where the first move is evaluated 1.50 “points” greater than the second best move according to the computer; see Study 3 for a full rationale behind this choice of criterion and a more detailed explanation). Problems satisfying the criterion of “tactical” included problems 1, 2, 8, 10, 11, and 15.

Tactical problems showed greater searching than positional problems for unique other moves, $t(156) = 1.61, p < .001$, mean depth of search, $t(156) = 2.85, p < .01$, and branchiness, $t(156) = 5.92, p < .001$. Positional problems showed greater searching than tactical problems for unique base moves, $t(156) = -6.27, p < .001$, episodes, $t(156) = -4.80, p < .001$, and null moves, $t(156) = -8.37, p < .001$. These data further suggest how search can be notably affected by idiosyncratic aspects of specific problems.

Finally, consider problem 12, which was the problem showing the greatest correlation between skill rating and problem accuracy as illustrated in Figure 2:



Figure 2: Problem 12 (left panel shows problem, right panel shows key aspects)

To understand this solution to this problem (1.d4-d5 or pawn to d5 square), consider the right panel of Figure 2. Notice that if this altered version were the position, white could move his queen to the diagonal highlighted by the red line, which would check black's king. Black would be unable to interpose either his queen or bishop as white could then capture either with impunity, so black must move his king instead, but only 2 squares are available (f8 and h8). If black moves his king to f8, then white plays Qf7 checkmate thanks to the supportive role of the knight on g5. On the other hand, if black plays the king to h8, then white can carry out a well-known maneuver that leads to a famous position known as the "smothered mate." Here is the sequence of moves that is "forced" (i.e. cannot be avoided) after 1.Qc4 Kh8, namely 2.Nf7 Kg8 3.Nh6 Kh8 4.Qg8 Rg8 5.Nf7 checkmate. Notice that this sequence alone is 9 ply deep, but the sequence of moves starting with 2.Nf7 can most likely be activated a single goal-generation production given that this sequence is quite common (as a unit) and almost always occurs in this manner (i.e. with these specific knight and black king maneuvers in this part of the board or its mirror image).

The question that can be immediately raised is whether this sequence alone produces the observed skill – depth of search correlation for this problem or whether the correlation would exist even for players who did not find this line. Fortunately, one final key aspect of the problem allows this to be investigated objectively. After the correct solution move 1.d5, black's most important response to consider is taking with the e-pawn, 1...exd5. Now white could play 2.Rd1 and put pressure on black's d-pawn (in exchange for his own d4 pawn), though black can defend with 2...Ne7 preventing white from immediately winning the pawn back. As I will discuss below, many very strong players opted for this idea. However, if white instead plays the very strong move 2.Bxd5, white gains a large advantage (after 2...Qxd5 3.Rd1 forces the queen off the highlighted diagonal shown in the right panel of Figure 2, allowing the forcing procedure mentioned above), but no player would play 2.Bxd5 unless they had understood the smothered mate possibility—otherwise, they would lose a bishop for essentially nothing in return. Hence, protocols showing 2.Bxd5 imply that white found both the correct solution move and the smothered mate idea, whereas protocols showing different second moves (e.g., 2.Rd1), indicate that this possibility was overlooked, even if the correct first solution move was discovered.

PRESTO assumes that search is based on discovered relationships. Considering the right panel of Figure 2, clear evidence for a goal-generation production emerges. The relationship <Ng5, Kg8, of7, oe6, od5, oc4, Qc2> (note that "o" represents an empty square) triggers goal positions of either <Ng5, Kf8, Qf7> or <Kh8, Rg8, Pg7, Ph7, Nf7>. To resolve this, the queen must be placed on the open diagonal. Hence, the player would engage in closure search for ways to put the queen on the diagonal and could also verbalize the resulting variation (illustrated previously).

For the whole sample, the Pearson correlation between rating and accuracy (choosing 1.d5) for this problem was $r(156) = .42$, $p < .001$, but it is interesting to examine whether the search correlations are driven by players who found the Bxd5

variation. Unfortunately, the raw protocol data for the Russian players in this sample ($N = 62$) was not available (protocol parameters were available and were analyzed in Study 1), so it is excluded in the following analyses (notably, there was no effect of study site on Elo rating, $p > .05$). For the restricted sample, the correlation between rating and accuracy (choosing 1.d5) for this problem was $r(95) = .31$, $p < .01$, and the significant correlations between rating and search statistics are as follows: accuracy, $r(95) = .31$, $p < .01$, unique other moves, $r(95) = .48$, $p < .001$, episodes, $r(95) = .32$, $p < .01$, max depth, $r(95) = .43$, $p < .001$, and mean depth, $r(95) = .32$, $p < .01$. Null moves, unique base moves, and branchiness were not significant, $p > .05$. By excluding these 9 players, I will examine whether the search correlations are maintained or whether identification of this single variation drives the correlations. In this final sample of 88 players, the significant correlations with skill were as follows: accuracy, $r(86) = .24$, $p < .05$, unique other moves, $r(86) = .45$, $p < .001$, episodes, $r(86) = .34$, $p < .001$, null moves, $r(86) = .21$, $p < .05$, max depth, $r(86) = .34$, $p < .001$, mean depth, $r(86) = .31$, $p < .01$. Unique base moves and branchiness were not significant, $p > .05$. Hence, removing players that found this idea does not change the correlations between search and skill, indicating that this idea alone (consisting of only a few relationships) did not drive these effects. This is not surprising since there are many other relationships in this position that can drive search, and this further suggests that the quantity of relationships, rather than a single relevant relationship, typically drives search correlations as argued by PRESTO.

Discussion

This study found evidence that the number of relationships in a position could predict whether better chess players search more deeply in that position. This is fully consistent with PRESTO, which argues that search revolves around discovered relevant relationships and that better players are better able to find such relationships due to acquired productions. Although little support was found for episodes – skill correlation predictability, this is most likely due to searching more episodes being essentially unimportant for accurately solving problems (with the exception of problem 10). Hence, a set of problems showing greater variability here ought to produce a significant correlation. Moreover, branchiness – skill correlations could not be predicted given that memory errors, rather than relationship discovery, are probably more relevant to this search parameter.

It should be noted that a null correlation between relationships and the depth – skill correlations would have been possible and reasonable a priori, particularly if deeper search was driven primarily by a specific relationship (i.e. relationship type) rather than the quantity of relationships. Therefore, the null hypothesis was not trivially refutable. Although some relationships are no doubt more conducive to search than others, the quantity of relationships is a key factor for predicting whether stronger

players will search more in a specific position as illustrated by the analysis of problem 12.

STUDY 3: HISTORIC INCREASES IN CHESS SKILL

Given that the PRESTO model equates chess skill with acquired productions, players with superior numbers and quality of productions will play superior chess, and this should hold true across chess history. This leads to the unique prediction that the best chess players of modern times are likely to play superior chess to the best players from earlier periods. This is because many factors are likely to lead modern players to have a superior set of productions.

First, modern players benefit from access to superior quality of information than earlier players. As chess literature evolves and new ways for evaluating structures and new types of goals and plans are discovered, modern players will be able to learn better quality structure-evaluation and goal-generation productions (the cumulative knowledge assumption).

Second, as the chess literature grows and becomes more widely available, the greater supply of chess instruction should lead to more potentially highly motivated individuals being able to get access to means for chess improvement.

Third, as chess tournaments and chess organizations have increased over time, more opportunities have arisen for a greater number of motivated individuals. This increases the range of individuals studying the game, including wider ranges on demographic variables, such as age. Hence, superior quality and quantity of practice is potentially available for modern players (see Howard, 2005 for an argument that the Flynn effect might lead to increased chess skill over time; notably, this argument has been severely criticized and is probably incorrect given that chess skill does not correlate with IQ).

Notably, an increase in chess skill at the highest levels is not guaranteed under alternative views, such as some notions of innate talent where the upper limit of achievable skilled performance is fixed by genes. In fact, some chess enthusiasts propose that the current world champion (at that time) was 100 rating points lower than the first world champion in 1886, based on similar statistical rating comparison (Clarke, 1963). Even former world champion Bobby Fischer, in his list of the top 10 players of all time, included five of the world champion contenders prior to 1935, and only two (out of a possible seven) after that date (Rowan, 2004).

Moreover, based on the notion of implicitly acquisition of chess chunks and templates, the highest level of skill at any point in time is a function of having learned the most chunks and templates, and from this an unchanging highest level could be expected. Given that cognitive learning parameters (e.g., time for encoding new information in LTM) have not changed in the last 150 years, the greatest number of chunks that can be learned by a human is unlikely to have changed in this time (notably

an assessment of chunk/template quality might lead to a prediction of historic increase, but template theory does not detail this aspect of chunks given its reliance on memory data). Clearly, an increase in chess ability over time is far from universally accepted.

Unfortunately, it is impossible to compare ratings of world champions to achieve this end—not only did the rating system not become widespread until the middle of the 20th century, but a player’s rating describes a player’s skill relative to competitors from the particular historic period. Comparing the ratings of one world champion to another is essentially meaningless since they played against different sets of potentially differentially skilled opponents. In contrast, the method proposed here uses powerful chess computer programs to measure the quality of each chess move in the archival games from world champion matches beginning in 1886 and ending in 2000. Moreover, as these records contain chess players’ actual games, they represent an ecologically valid assessment of players’ performances at a given point in time.

I used one of the most powerful chess computer programs to objectively evaluate these recorded moves. Computer chess programs can defeat virtually all human players by using “brute force” search algorithms, though they only search a limited depth and cannot find “perfect” moves. Nonetheless, within the scope of their search, computer programs are virtually perfect at detecting *tactical errors*. That is, they can discover any inferior move that leads to large losses or any superior move that eventually results in large gains, these being measured in terms of a material advantage in which one player has more chess pieces or more valuable pieces than the other.

If computer-assessed tactical-error rates are an objective and reliable measure of chess skill, the tactical error rates from the matches for world chess champion can be compared across historic time. Moreover, data from players’ biographies could be used to extract relevant demographic information to potentially explain any historic effects. According to PRESTO, superior practice histories should explain a rise in chess skill at the highest levels of performance. Modern chess champions should have a greater number and superior quality of productions in LTM, playing better chess by making fewer mistakes.

Method

Participants

The dataset from study 1 and 2 will be used here to address correlations between biographic variables (see below) and metrics of chess practice and skill.

Materials

Using a commercially available database from ChessBase (ChessBase, Hamburg, Germany), I generated the set of the 802 games from the chess world championships. This included all championship matches between two competitors from 1886 to 2000. I used the *Fritz 8.0* (ChessBase, Hamburg, Germany) chess playing computer program to analyze the database on a Gateway 2.4 GHz processor with 512 MB of RAM. *Fritz 8.0* was set to “Blunder Check” mode with a fixed depth of 12 ply and all other settings remaining at their defaults.

For validity and reliability calculations, I also generated a set of games from recent tournaments with rated chess players. These games were selected from three sources to obtain a wide range of ratings. The first source included two tournaments from the ChessBase database, 90 games from Hoogovens, 2003 (Elo 2,100 to 2,600) and 16 games from Astana, 2001 (2,700+). The second source was a collection of 43 class player (Elo 1,000 to 2,200) games from USCF-rated tournaments. These games were selected based on a criterion that players be separated by no more than 100 rating points in order to approximate the conditions of the world championships. Finally, I created a data set for each rating range of 100 points, from 1,100 to 2,800, taking the average of each player’s tactical-error rate within each category to examine this rate across each of the 17 categories of ratings.

I also obtained data on the development of each world champion contender from biographical sources. I established each player’s average age in world championships (“average age”), the starting age when they learned the chess rules (“starting age”), the age when they became serious about studying chess (“serious age”), as well as additional information, such as whether they learned the rules from parents (“parental instruction”) and whether they had another occupation besides professional chess (“another occupation”).

Notably, biographical sources could produce little information on practice activities. I derived the serious age based on the age directly indicated by the player. In absence of this I used other indicators of serious commitment, such as joining a chess club or beginning to study chess books. In some cases, the specific age could only be narrowed to a small range, and we chose the mean age of that interval. The excerpts and exact references to biographic sources are available from the author.

Procedure

The *Fritz 8.0* program analyzed all moves of all games in the world championship set. After every move, *Fritz* searched a depth of 12 ply (6 moves), recording its evaluation after each move using a standard metric for assessing the superiority/inferiority of white’s position in terms of pawns. For example, +5.0 indicates white is winning by roughly a 5 pawn advantage, and -0.33 would indicate black is slightly ahead by a fraction of a pawn. I use the same operational definition of tactical errors proposed recently by Chabris and Hearst (2003)—this definition of errors is based on the difference in the computer’s evaluation between consecutive moves being above a specific threshold, namely 1.50 pawns.

Moreover, if the previous position's evaluation favored one side by 3.0 pawns or more and the position changed by 1.50 after a chosen move but remained evaluated above 3.0 pawns for the same side, the move was excluded as a tactical error because virtually all games are still considered technically won at that point. In fact, most players in this database resigned when the position reached this point.

Results

As an initial validation of our measure of tactical errors for all of the games in the 37 championship matches, I compared the difference in the number of tactical errors committed by the player of white to that of the player of black with the outcome of the corresponding game. I restricted the analysis to games where one player committed at least one more tactical error than the other. I found a significant correlation between tactical errors and outcome for white, $r(197) = -.82, p < .001$, when the draws were excluded the correlation decreased to $r(169) = -.89, p < .001$ (28/199 games). I examined the split-half reliability of tactical error rates within individuals in this world championship sample. For each of the 26 competitors, we calculated their tactical error rate based on the accumulations from the odd numbered and from the even numbered games. The correlation between even and odd number based error rates was high, $r(24) = .70, p < .001$, especially for such a highly selected sample of world-class players.

To validate that tactical-error rates measure chess skill, I examined the relationship between error rates and chess ratings in our sample of 17 groups (i.e., of contemporary players with similar chess ratings, such as between 1800 and 1899), $r(15) = -.81, p < .001$. At the highest and lowest categories of chess skill we had difficulties finding many data points. Consequently, I excluded three categories with only a single data point each, and found an even stronger correlation, $r(12) = -.90, p < .001$, though there was no evidence for quadratic effects, $p > .05$. Notably, the split-half reliability estimate for these tactical-error rates was $r(5) = .87, p < .05$. Moreover, I found additional evidence supporting lower rates of tactical errors for more skilled players. Games where neither player made tactical errors would be more likely to result in a draw. The percentage of drawn games in each rating group increased with the chess rating of the chess players, $r(12) = .76, p < .01$, as would be expected from having lower tactical-error rates. Importantly, in this analysis I excluded all games ending before move twenty, as many strong grandmasters agree to draws with equally strong opponents after only playing a few memorized opening moves (memorized opening variations often end before move twenty). Therefore, I found support for the hypothesis that the measure of tactical error rate is a reliable and valid indicator of chess skill.

I calculated average rates of tactical errors for each of the 37 world championship matches and this correlated with the year of the championship, $r(35) = -.62, p < .001$. Several chess champions played in multiple championships, and we also calculated average tactical-error rates across all of their matches for each of the 26 unique players. These average error rates were closely associated with historic year, $r(24) = -.68, p <$

0.001, and this is illustrated in Figure 2. Notably, these estimates of tactical-error rates for more modern champions are consistent with those reported for groups of contemporary grandmasters by other researchers (e.g., Chabris & Hearst, 2003).

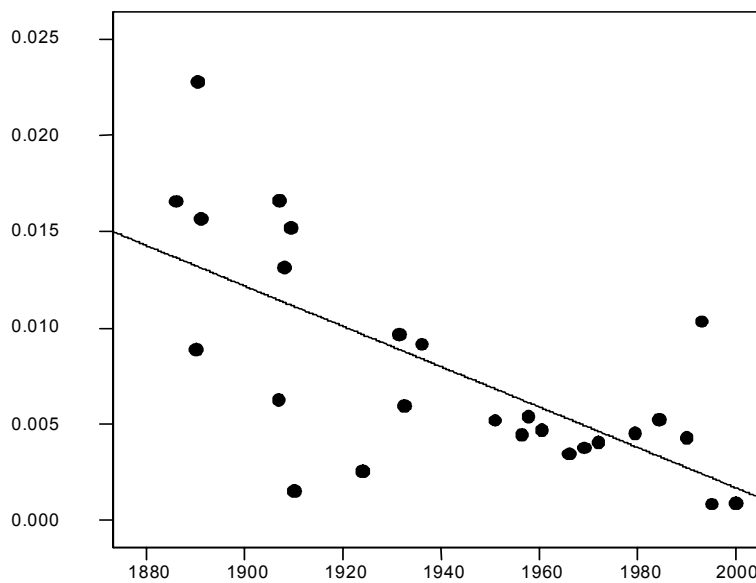


Figure 3: Declining Tactical Error Rates of World Champions. Tactical Error Rate as a Function of Historic Year.

Given that tactical-error rates strongly predict chess skill, these results imply dramatic improvements at the highest levels of intellectual achievement in the game of chess over the last two centuries. Similar to the finding that draws are more frequent in higher rated groups, I found that the frequency of drawn games in each match has also increased historically, $r(35) = .70, p < .001$. Further, the process of recruiting a challenger for the world championships only became officially regulated around 1948, and earlier champions had more control over who could challenge them for the title. Thus, I restricted the analysis to the world champions (omitting players who never won the title) and replicated the negative correlation $r(12) = -.67, p < 0.01$, showing that the historic improvements cannot be explained by early world champions selecting weaker opposition. Notably, fitting nonlinear functions to the data in Figure 3 did not explain additional reliable variance, supporting the idea that chess skill has continued to improve linearly throughout the century.

To understand this historic improvement in chess ability, I conducted an analysis of biographic variables. Two of the five variables, learning from parents and having another occupation, did not correlate reliably with either year or error rate. However, the average age and starting age reliably correlated with year and with error rate. Finally, serious age revealed the highest correlations with year, $r(23) = -.76, p < .001$, and with error rate, $r(23) = .70, p < .001$, shown in Figure 3. Note that serious age was could not be obtained for one of our players, David Janowski, resulting in one less degree of freedom.

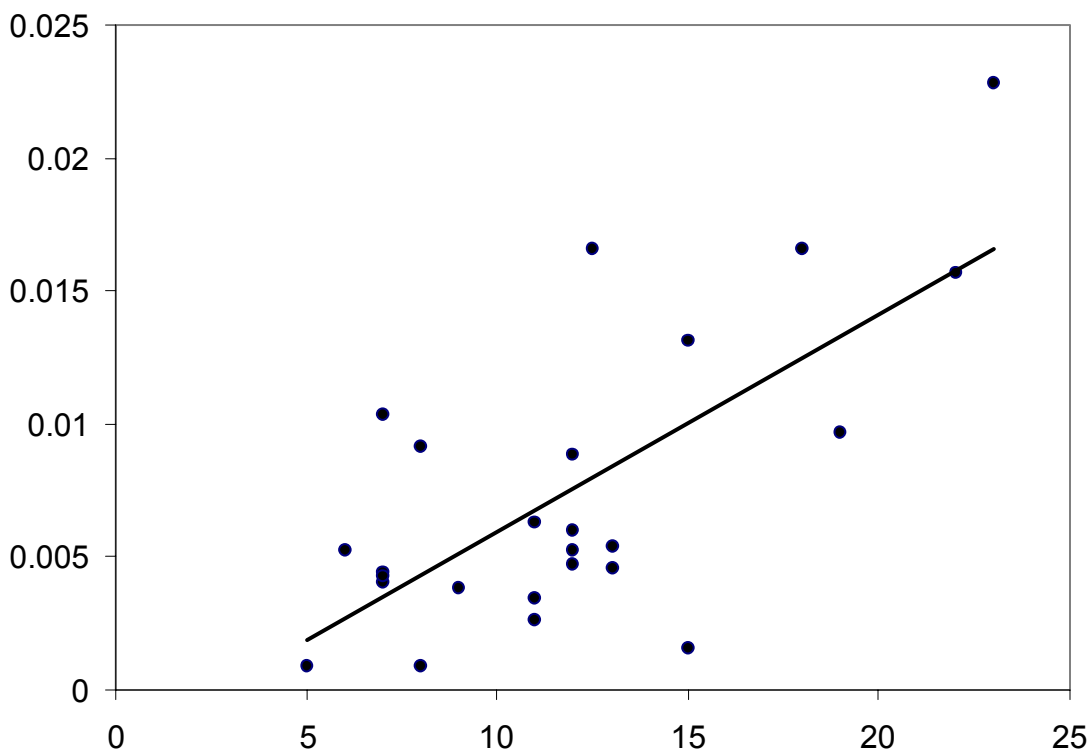


Figure 4: Serious Age Varies (x-axis) with Tactical Error Rate

I tested whether serious age could account for all of our findings. When I entered the other four biographical variables into a regression equation predicting tactical error rate, this first model explained reliable variance, $R^2_{change} = .45, F(4, 18) = 3.731, p < .05$ (Note that there was no information on whether 2 of the players learned from parents, leading to 18 degrees of freedom). However, next entering serious age explained additional reliable variance, $R^2_{change} = .15, F(1, 17) = 6.307, p < .05$, (the individual without serious age information was excluded, further reducing the degrees of freedom). I then entered historic year into the equation, but this did not explain additional variance in error rate, $p > .05$, and

finally I removed the original four biographic variables, which did not reliably reduce variance, $p > .05$.

In the final model neither year, $p > .05$, nor serious age, $p > .05$, uniquely predicted reliable variance in tactical errors. Therefore, serious age fully accounted for the historic variance in tactical error rate and was statistically superior to all other biographic variables. Notably, a regression analysis with only serious age and year predicting error rate leads to the same result and thus shows the robustness of the results.

Finally, I examined whether serious age predicts chess rating and accumulated solitary practice in the original experimental sample. As reported in Charness et al. (2005), serious age does not correlate with total accumulated solitary practice (see also Gobet & Campitelli, 2007), though it does correlate with rating. However, an inspection of the scatterplot reveals a relationship, as shown in Figure 5:

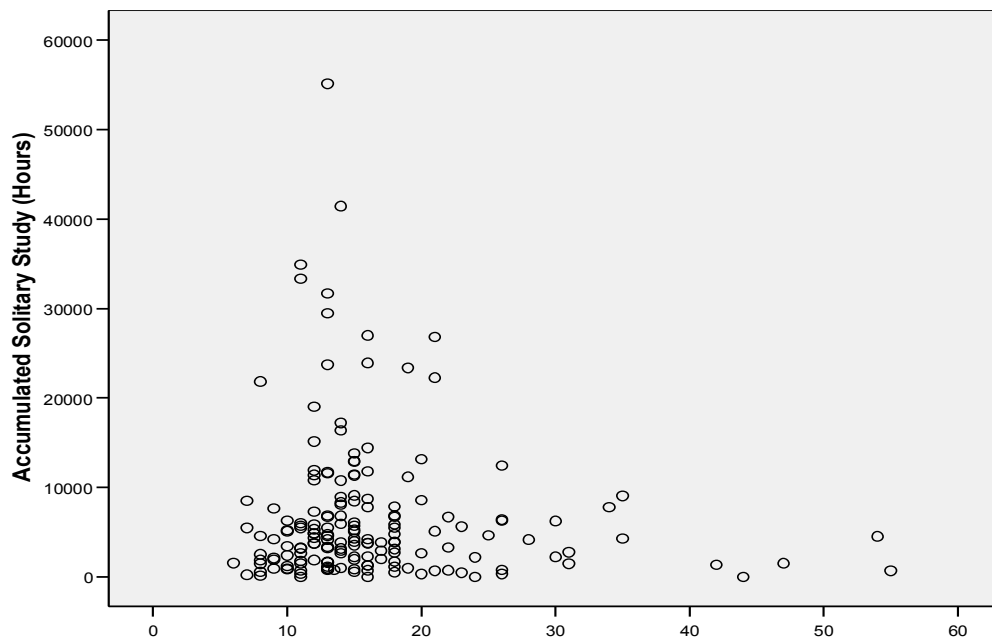


Figure 5: Accumulated Solitary Study as a Function of Serious Age (x-axis)

From the graph it appears that heteroskedasticity obscures the relationship, given that no participant over age 25 accumulates over 20,000 reported hours of practice—not all players who began at young ages engaged in large amounts of study. To analyze this differently, I first calculated, for each serious age, the proportion of sample participants acquiring at least 5,000 hours of solitary practice, approximately the amount needed to acquire expert-performance levels from previous studies (Charness et al., 2005). The correlation was significant, $r(29) = -.49$, $p < .001$, indicating that initiating serious chess

study at later ages reduces the likelihood of gaining large amounts of practice. Serious age was also correlated with chronological age, $r(173) = .37, p < .001$, indicating that more recent cohorts tended to initiate serious chess study at younger ages, supporting the general finding.

Discussion

This study provides strong evidence that the highest levels of chess skill have risen over the last two centuries. Modern world champions commit fewer tactical errors than their predecessors, and this is explained by their starting serious chess study at younger ages. Comparing the initial age of serious chess study to training from the data in study 1, I found evidence that starting at younger ages tended to predict much greater amounts of solitary study, though the relationship is non-linear. The greatest amounts of solitary study were accumulated by players starting serious study as early as age 10 and the advantage from early starts tended to level off as players approached adulthood around age 20. This suggests that greater free time, fewer obligations, and other superior factors of the childhood environment can allow greater amounts of practice to accumulate. Hence, as predicted by PRESTO, superior practice is linked to superior play, even for the top players in the world.

GENERAL DISCUSSION

A new theory of chess skill has been proposed in this manuscript that centers on the explicit acquisition of productions to facilitate the discovery, temporary storage, evaluation, and generation of relationships between chess pieces. This is the first cognitive psychological theory of chess move-selection and search behavior and it relates to broader frameworks from cognitive psychology. It improves over older models of chess skill, such as chunking and template theory, in its greater consistency with a host of findings, particularly the unimportance of mere experience for skill improvement, and in its ability to make novel predictions regarding the search characteristics of stronger players and the unique finding of a historic rise in chess skill.

While this model has been primarily successful across the three empirical studies examined here, it should be considered only a first step to fully explaining individual differences in chess skill. For instance, more work is needed to determine the precise manner in which new productions are explicitly acquired and ultimately consolidated in memory. Computational approaches can also extend the present work and with greater understanding and incorporation of perceptual and memory processes, computer programs could be developed to play chess as humans do, rather than merely play chess as well as or better than humans.

Finally, PRESTO's core idea of increasing specificity of acquired productions during skill development may help researchers understand skill acquisition in many other domains. To what extent the aspiring performer must move beyond applying

general principles to applying highly specific principles remains to be determined, but many domains may be compatible with this view. In chess play, the performer must encode situation-specific information rapidly, evaluate it, and produce possible solution ideas, compatible with the three types of productions. Other domains may be similarly described, since many depend on cognitive intake of situation-specific information. This may include not only other intellectual games, such as bridge, GO, or checkers, but also potentially many other superficially different domains, such as second language acquisition and perceptual athletic skills. Only future research can ultimately determine this.

Overall, the present work incorporates and illustrates the techniques and methodology of the Expert-Performance Approach (cf. Ericsson & Smith, 1991), and combines process tracing techniques (viz. protocol analysis) to understanding tasks representative of domain skill (viz. chess problems) in tandem with practice questionnaires and diaries to determine how the mechanisms mediating superior performance (viz. productions) are acquired (viz. deliberate practice). This approach is being applied to many other domains will success and will continue to serve as a model method for capturing real-world skill in the psychologist's laboratory.

Chess is one of the oldest cognitive activities and today is enjoyed by players in many different countries around the world. Whether a single framework such as that proposed here truly captures the fundamental complexities of countless minds attempting to make outstanding chess moves is difficult to know with any certainty. If nothing else, the present work aims to make progress on this challenging problem and to suggest a starting point for grappling with the acquisition and demonstration of a complex skill.

APPENDIX A: INFORMED CONSENT FORM

INFORMED CONSENT FORM

Date: _____

I, _____, freely and voluntarily and without undue inducement or any element of force, fraud, deceit, duress, or other form of constraint or coercion, consent to be a participant in the research project entitled "Life span Expertise" to be conducted at Tallahassee during the period 200-2002, with Neil Charness, Jeffrey Feddon and Michael Tuffiash as Principal Investigators. The procedures to be followed, and their purposes, including identification of any procedures which are experimental, have been explained to me and I understand them. They are as follows: To fill out questionnaires about chess practicing behavior, participate in a computer session (~1.5 hrs) that will involve reconstructing board positions and choosing the best move in a chess game, and keep a record of chess practice and preparation for the week following the computer session.

The attendant discomforts and risks reasonably to be expected by my participation in subject project have been explained to me and I understand the to be as follows: There are no significant risks. Since the sessions are sometimes longer than an hour, there may be some fatigue, though rest breaks will be allowed.

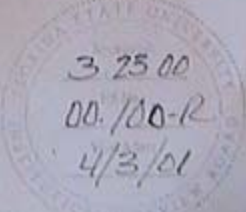

Any benefits reasonably to be expected from my participation and any alternative procedures that might be advantageous have been explained to me and are as follows: We expect to show skill as a predictor of performance due to differences in information grouping.

The records of this research which identify you will be afforded the following confidentiality. They will be kept in locked files, with names separate from identification numbers. Any publication of results will deal with group data so that your identity is not revealed. The master link will be destroyed once the study is completed by 2005.

I understand that this consent may be withdrawn at any time without prejudice, penalty or loss of benefits to which I am otherwise entitled. I have been given the right to ask and have answered any inquiry concerning the foregoing. Question, if any, have been answered to my satisfaction. In the future, I understand I may contact: Neil Charness, Jeffrey Feddon, or Michael Tuffiash, Department of Psychology, Florida State University, Tallahassee, Florida, USA 32306-1270, Phone 850-644-6686 for answers to pertinent questions about this research, my rights, or in case of a research-related injury. I have read and understand the foregoing.

Subject _____

DATE _____



The left stamp is circular with the text "FLORIDA STATE UNIVERSITY" at the top and "INSTITUTIONAL REVIEW BOARD" at the bottom. Inside the circle, it says "Approved 3/16/01", "IRB No. 01-147R", and "Valid Through 4/3/02". The right stamp is also circular with the same text. Inside, it says "3 25 00", "00-100-R", and "4/3/01".

APPENDIX B: IRB APPROVAL LETTER



Office of the Vice President
for Research
Tallahassee, Florida 32306-2763
(850) 644-8633 • FAX (850) 644-4392

REAPPROVAL MEMORANDUM

from the Human Subjects Committee

Date: January 30, 2001

From: David Quadagno, Chairperson *DQ/ph*

To: Neil Charness

MC: 1270

Dept: Psychology

Re: Reapproval of Use of Human subjects in Research

Project entitled: Life-Span Expertise

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 April 2, 2002 please request renewed approval.

You are reminded that a change in protocol in this project must be approved by resubmission of the project to the Committee for approval. Also, the principal investigator must report to the Chair promptly, and in writing, any unanticipated problems involving risks to subjects or others.

By copy of this memorandum, the Chairman 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 of such investigations as often as necessary to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

:hh

CC:

human/renewal.hs

APPLICATION NO. 01147 -R

REFERENCES

- Ackerman, P.L., Beier, M.E., & Boyle, M.O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131, 30-60.
- Allard, F., Graham, S., & Paarsalu, M. E. (1980). Perception in sport: Basketball. *Journal of Sport Psychology*, 2, 14-21.
- Amidzic, O., Riehle, H. J., Fehr, T., Wienbruch, C., & Elbert, T. (2001). Pattern of focal gamma bursts in chess players. *Nature*, 412, 603.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press.
- Baddeley, A. (1986). *Working Memory*. Oxford University Press.
- Barbaranelli, C., Caprara, G.V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire measuring the Big Five in late childhood. *Personality and Individual Differences*, 34, 645-664.
- Bartlett, F.C. (1932). *Remembering*. Cambridge University Press: Cambridge.
- Baumgarten, F. (1930). *Wunderkinder, psychologische Untersuchungen*. Leipzig: Johann Ambrosius Barth.
- Bilalič, M. (2006) Acquisition of chess skill. Unpublished doctoral dissertation, Oxford University, Oxford, United Kingdom.
- Bilalič, M., McCleod, P., & Gobet, F. (2007a). Does chess need intelligence? A study with young chess players. *Intelligence*, 35, 457-470.
- Bilalič, M., McCleod, P., & Gobet, F. (2007b). Personality profiles of young chess players. *Personality and Individual Differences*, 42, 901-910.
- Binet, A. (1894). *Psychologie des Grands Calculateurs et Joueurs d'Echecs*. Paris: Librairie Hachette.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *The Quarterly Journal of Experimental Psychology*, 10, 12-21.

- Bryan, W. L., & Harter, N. (1899). Studies in the telegraphic language. The acquisition of a hierarchy of habits. *Psychological Review*, 6, 345-375.
- Bugelski, B. R. (1962). Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology*, 63, 409-412.
- Burns, B. D. (2004). The effects of speed on skilled chess performance. *Psychological Science*, 15, 442-447.
- Campitelli, G., & Gobet, F. (2004). Adaptive expert decision making: Skilled chess players search more and deeper. *ICGA Journal*, 209-216.
- Campitelli, G., Gobet, F., Head, K., Buckely, M., & Parker, A. (2007). Brain localisation of memory chunks in chessplayers. *International Journal of Neuroscience*, 117, 1641-1659.
- Campitelli, G., Gobet, F., Williams, G., & Parker, A. (2007). Integration of perceptual input and visual imagery in chess players: Evidence from eye movements. *Swiss Journal of Psychology*, 66, 201-213.
- Chabris, C.F. & Hearst, E. S. (2003). Visualization, pattern recognition, and forward search: Effects of playing speed and sight of the position on grandmaster chess errors. *Cognitive Science*, 27, 637-648.
- Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 641-653.
- Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology*, 33, 1-16.
- Charness, N. (1981a). Aging and skilled problem solving. *Journal of Experimental Psychology: General*, 110, 21-38.
- Charness, N. (1981b). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 467-476.
- Charness, N. (1981c). Visual short-term memory and aging in chess players. *Journal of Gerontology*, 36, 615-619.
- Charness, N. (1989). Expertise in chess and bridge. In D. Klar & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp.183-208).

Hillsdale, NJ: Erlbaum.

Charness, N. (1992). The impact of chess research on cognitive science. *Psychological Research*, 54, 4-9.

Charness, N., Krampe, R., & Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: An international comparison of life-span chess skill acquisition. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the Arts and Sciences, Sports and Games* (pp. 51-80). Mahwah, NJ: Erlbaum.

Charness, N., Reingold, E.M., Pomplun, M., & Stampe, D.M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition*, 29, 1146-1152.

Charness, N., Tuffiash, M., & Jastrzemski, T. (2004). Motivation, emotion, and expert skill acquisition. In D. Dai & R. J. Sternberg (Eds.) *Motivation, emotion, and cognition: Integrative perspectives* (pp. 299-319). Mahwah, NJ: Erlbaum.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E. M., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, 19, 151-165.

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

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

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp.73-96). Hillsdale, NJ: Erlbaum.

Cleveland, A. A. (1907). The psychology of chess and learning to play it. *American Journal of Psychology*, 18, 269-308.

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

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-114.

- de Bruin, A.B.H., Rikers, R.M.J.P., & Schmidt, H.G. (2007). The influence of achievement motivation and chess-specific motivation on deliberate practice. *Journal of Sport and Exercise Psychology, 29*, 561-583.
- de Groot, A. D. (1946/1965/1978). *Thought and choice in chess*. The Hague: Mouton.
- de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess*. Assen (The Netherlands): Van Gorcum.
- Der Spiegel (1987). Genieblitze und Blackouts. Issue 52187 (p. 126-140).
- Djakow, I. N., Petrowski, N. W., & Rudik, P. A. (1927). *Psychologie des Schachspiels*. Berlin: de Gruyter.
- Doll, J., & Mayr, U. (1987). Intelligenz und schachleistung - eine untersuchung an schachexperten. *Psychologische Beiträge, 29*, 270-289. (Intelligence and chess performance - An investigation of chess experts.)
- Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition, 7*, 149-158.
- Elo, A. E. (1965). Age changes in master chess performances. *Journal of Gerontology, 20*, 289-299.
- Elo, A. E. (1986). *The rating of chessplayers, past and present*, (2nd ed.). New York: Arco.
- Engle, R. W., & Bukstel, L. (1978). Memory processes among bridge players of differing expertise. *American Journal of Psychology, 91*, 673-689.
- Ericsson, K. A., & Harris, M. S. (1990). *Expert chess memory without chess knowledge: A training study*. Poster presented at the 31st Annual Meeting of the Psychonomic Society, New Orleans, Louisiana.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.
- Ericsson, K. A., & Kintsch, W. (2000). Shortcomings of generic retrieval structures with slots of the type that Gobet (1993) proposed and modeled. *British Journal of Psychology, 91*, 571-590.
- Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice

- in the acquisition of expert performance. *Psychological Review*, 100, 363-406.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of adaptation to task constraints. *Annual Review of Psychology*, 47, 273-305.
- Ericsson, K. A., Oliver, W. L. (1988). In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought* (pp. 392-428). New York, NY, US: Cambridge University Press.
- Ericsson, K. A., & Roring, R. W. (2007). Memory use as skilled cognition. *The Psychology of Learning and Motivation*, 48, 351-380.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.
- Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.) *Studies of Expertise: Prospects and limits* (pp. 1-38). New York: Cambridge University Press.
- Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235-267). Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Feigenbaum, E. A., & Simon, H. A. (1962). A theory of the serial position effect. *British Journal of Psychology*, 53, 307-320.
- Feldman, D. H., & Katzir, T. (1998). Natural talents: an argument for the extremes. *Behavioral and Brain Sciences*, 21, 414-415.
- Frey, P. W., & Adesman, P. (1976). Recall memory for visually presented positions. *Memory and Cognition*, 4, 541-547.
- Frydman, M., & Lynn, R. (1992). The general intelligence and spatial abilities of gifted young Belgian chess players. *British Journal of Psychology*, 83, 233-235.
- Gobet, F. (1997). Roles of pattern recognition and search in expert problem solving. *Thinking and Reasoning*, 3, 291-313.
- Gobet, F. (1998). Expert memory: a comparison of four theories, *Cognition*, 66, 115-

152.

- Gobet, F. (2000a). Some shortcomings of long-term working memory. *British Journal of Psychology*, *91*, 551-570.
- Gobet, F. (2000b). Retrieval structures and schemata: A brief reply to Ericsson and Kintsch. *British Journal of Psychology*, *91*, 591-594.
- Gobet, F., & Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology*, *43*, 159-172.
- Gobet, F., & Charness, N. (2006). Expertise in chess. In K. A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.) *Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four...or is it two? *Memory*, *12*, 732-747.
- Gobet, F., & Jackson, S. (2002). In search of templates. *Cognitive Systems Research*, *35-44*.
- Gobet, F., & Jansen, P. (1994). Towards a chess program based on a model of human memory. In H.J. van den Herik, I.S. Herschberg, & J.W. Uiterwijk (Eds.), *Advances in Computer Chess 7*. Maastricht: University of Limburg Press.
- Gobet, F. & Simon, H. A. (1996a). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*, 1-40.
- Gobet, F., & Simon, H. A. (1996b). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, *3*, 159-163.
- Gobet, F., & Simon, H. A. (1996c). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from Grandmaster level chess. *Psychological Science*, *7*, 52-55.
- Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, *6*, 225-255.
- Gobet, F. & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, *24*, 651-682.
- Gold, A., & Opwis, K. (1992). Methods for the empirical analysis of chunks in recalling

- chess positions. *Sprache and Kognition*, 11, 1-13.
- Goldin, S. E. (1979). Recognition memory for chess positions: Some preliminary research. *American Journal of Psychology*, 92, 19-31.
- Grabner, R. H., Neubauer, A. C. & Stern, E. (2006) Superior performance and neural efficiency: The impact of intelligence and expertise. *Brain Research Bulletin* 69:422-39.
- Grabner, R. H., Stern, E., & Neubauer, A.C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, 124, 398-420.
- Holding, D. H. (1979). The evaluation of chess positions. *Simulation and Games*, 10, 207-221.
- Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Holding, D. H. (1992). Theories of chess skill. *Psychological Research*, 54, 10-16.
- Holding, D. H., & Pfau, H. D. (1985). Thinking ahead in chess. *American Journal of Psychology*, 98, 271-282.
- Holding, D. H., & Reynolds, R.I. (1982). Recall or evaluation of chess positions as determinants of chess skill. *Memory and Cognition*, 10, 237-242.
- Horgan, D. D., & Morgan, D. (1990). Chess expertise in children. *Applied Cognitive Psychology*, 4, 109-128.
- Howard, R. W. (1999). Preliminary real-world evidence that average human intelligence is rising. *Intelligence*, 27, 235-250. Howard, R. W. (2001). Searching the real world for signs of rising population intelligence. *Personality and Individual Differences*, 30, 1039-1058.
- Howard, R. W. (2005). Objective evidence of rising population ability: A detailed examination of longitudinal chess data. *Personality and Individual Differences*, 38,347-363.
- Hulin, C. L., Henry, R. A., & Noon, S. Z. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin*, 107, 328-340.

- Krogius, N. (1976). *Psychology in chess*. New York: RHM Press.
- Lane, D. M., & Robertson, L. (1979). The generality of the levels of processing hypothesis: An application to memory for chess positions. *Memory and Cognition*, 7, 253-256.
- Lassiter, G. D. (2000). The relative contributions of recognition and search-evaluation processes to high level chess performance: Comment on Gobet and Simon. *Psychological Science*, 11, 172-173.
- Lories, G. (1987). Recall of random and non random chess positions in strong and weak chess players. *Psychologica Belgica*, 27(2), 153-159.
- McGregor, S.J., & Howes, A. (2002). The role of attacks and defense semantic in skilled players' memory for chess positions. *Memory & Cognition*, 30, 707-717.
- McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, 13, 307-325.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Minsky, M. (1977). Frame-system theory. In P.N. Johnson-Laird and P.C. Wason, *Thinking. Readings in Cognitive Science*. Cambridge: Cambridge University Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Nimzovitch, A. (1930). *My System*. Harcourt, Brace.
- Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193-198.
- Pfau, H. D., & Murphy, M. D. (1988). Role of verbal knowledge in chess skill. *American Journal of Psychology*, 101, 73-86.
- Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science*, 12, 48-55.

- Reingold, E. M., Charness, N., Schultetus, R. S., & Stampe, D. M. (2001). Perceptual automaticity in expert chess players: Parallel encoding of chess relations. *Psychonomic Bulletin and Review*, 8, 504-510.
- Reitman, J.S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology*, 8, 336-356.
- Reynolds, R.I. (1982). Search heuristics of chess players of different calibres. *American Journal of Psychology*, 95, 383-392.
- Robbins, T.W., Anderson, E.J., Barker, D.R., Bradley, A.C., Fearnough, C., Henson, R., Hudson, S.R., & Baddeley, A.D. (1996). Working memory in chess. *Memory & Cognition*, 24, 83-93.
- Roring, R.W., & Charness, N. (2007). A multilevel model analysis of expertise in chess across the lifespan. *Psychology and Aging*, 22, 291-299.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.
- Saariluoma, P. (1985). Chess players' intake of task-relevant cues. *Memory and Cognition*, 13, 385-391.
- Saariluoma, P. (1989). Chess players' recall of auditorally presented chess positions. *European Journal of Psychology*, 1, 309-320.
- Saariluoma, P. (1990a). Apperception and restructuring in chess players' problem solving. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, and G. Erdos (Eds.). *Lines of thinking: Reflections on the psychology of thought, volume 2.* (pp. 41-57). London: John Wiley.
- Saariluoma, P. (1990b). Chess players' search for task relevant cues: Are chunks relevant? In D. Brogan (Ed.) *Visual search* (pp. 115-121). London: Taylor & Francis.
- Saariluoma, P. (1992a). Error in chess: The apperception-restructuring view. *Psychological Research*, 54, 17-26.
- Saariluoma, P. (1992b). Visuospatial and articulatory interference in chess players' information intake. *Applied cognitive psychology*, 6, 77-89.
- Saariluoma, P. (1994). Location coding in chess. *Quarterly Journal of Experimental*

- Psychology*, 47A, 607-630.
- Saariluoma, P. (1995). *Chess players' thinking: a cognitive approach*. London: Routledge.
- Saariluoma, P., & Laine, T. (2001). Novice construction of chess memory. *Scandinavian Journal of Psychology*, 42, 137-146.
- Schultetus, S. & Charness, N. (1999). Recall vs. position evaluation revisited: The importance of position-specific memory in chess skill. *American Journal of Psychology*, 112 (4), 555-569.
- Simon, H. A. (1974). How big is a chunk? *Science*, 183, 482-488.
- Simon, H. A., & Barenfeld, M. (1969). Information processing analysis of perceptual processes in problem solving. *Psychological Review*, 76, 473-483.
- Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, 61(4), 394-403.
- Simon, H. A., & Gilmarin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, 5, 29-46.
- Sloboda, J. A. (1976). Visual perception of musical notation: Registering pitch symbols in memory. *Quarterly Journal of Experimental Psychology*, 28, 1-16.
- Unterrainer, J. M., Kaller, C. P., Halsband, U., & Rahm, B. (2006). Planning abilities and chess: A comparison of chess and non-chess players on the Tower of London task. *British Journal of Psychology*, 97, 299-313.
- 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.
- Wagner, D. A., & Scurrah, M. J. (1971). Some characteristics of human problem-solving in chess. *Cognitive Psychology*, 2, 454-478.
- Waters, A. J., Gobet, F., & Leyden, G. (2002). Visuospatial abilities of chess players. *British Journal of Psychology*, 93, 557-565.

Watson, J. (1998). *Secrets of Modern Chess Strategy*. Gambit Publications.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, 72, 89-94.

BIOGRAPHICAL SKETCH

Roy Winn Roring III (Tres) was born and raised in Oklahoma. He earned a B.S. in computer science with minors in mathematics and physics at the University of Oklahoma in 2003 and an M.S. in cognitive psychology at the Florida State University in 2005. He has authored papers published in psychological journals and has presented at international conferences. He remains fascinated with individual differences and hopes to continue exploring this topic across many domains. He currently plans to take a postdoctoral position at the University of Sydney in Australia to apply psychological understanding of expert GO players to developing realistic and high performing artificial intelligence.