Cognitive and Neuropsychological Mechanisms of Expertise: Studies with Chess Masters

A thesis presented

by

Christopher Fitzgerald Chabris

to

The Department of Psychology

in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of

Psychology

Harvard University
Cambridge, Massachusetts

May 1999

Abstract

Research on expertise in visual-spatial tasks such as chess has ignored the introspective reports of top-level practitioners, overemphasized pattern recognition as the sole mechanism underlying skilled performance, and neglected the role of mental imagery in the thinking process. In addressing these limitations I propose a new theory of expertise, the “mental cartoons hypothesis,” and illustrate its properties by application to chess, which has historically been the primary domain for psychological studies of human expertise.

Experiment 1 uses a classic image-scanning paradigm to show that chess masters and novices differ substantially in their ability to visualize chess moves, even in semantically impoverished contexts, extending the range of novice-expert differences from pattern recognition and knowledge representation to mental imagery processing. Experiments 2 and 3 exploit original very-long-term memory recognition and recall tasks to show that the memory representations of famous chess positions held by chess masters include both pattern and conceptual information, supporting a key distinction between this theory and its predecessors. Experiment 4 uses computer analysis of 1188 chess games between grandmasters to show that when players have additional time to think ahead, the quality of their decisions improves significantly, refuting a claim extrapolated from theories that emphasize fast pattern recognition over slow search processes.

Experiments 5 and 6 investigate hemispheric specialization for chess perception, an issue not addressed by previous theories, and find that chess masters have a right-hemisphere advantage for recognizing previously studied normal chess positions and for

parsing normal chess positions into component patterns. This is consistent with other neuropsychological evidence that ties the right hemisphere to chess skill.

I also discuss more specific brain mechanisms that may support chess expertise, compare the mental cartoons hypothesis to other theories, suggest several specific experiments and general directions for future research, and argue that the study of expertise is relevant to a broad range of issues in human cognition.

Contents

Introduction

Chess is the oldest game in history, the one practiced most widely throughout the world, and the quintessential mental sport. According to Goethe, “chess is the touchstone of the intellect,” and researchers in artificial intelligence believed that “if one could devise a successful chess machine, one would seem to have penetrated to the core of human intellectual endeavor” (Newell, Shaw, & Simon, 1958). Although such claims are probably exaggerated, it is clear that research on how human beings play chess can be fruitful for psychology, especially when it addresses the broader issue of how exceptional performance, or expertise, differs from normal or novice performance.

Indeed, chess has been the primary arena for studies of expertise throughout this century. Among the advantages of chess as a research domain are its external and ecological validity, the broad range of abilities and tasks that can be studied, the availability of archives of historical records of expert performances, the ability to compare human performance with computational cognitive models and artificial intelligence research programs, and the statistical rating system that enables precise comparisons of skill between players (see Gobet, 1998b for further discussion). All of these advantages will be exploited in the theory and experiments presented in this thesis.

The importance of studying expertise

Theories of expertise are important for at least three reasons. First, the logic of psychological research emphasizes controlled comparison—the manipulation of a single variable whose effects can be closely monitored. To understand how a system of interacting processing components works, it would be useful to selectively degrade the performance of some components and observe the effects on the overall system; this is the basic reason for studying brain-damaged and psychiatric patients, for using transcranial magnetic stimulation to create “virtual brain lesions” (e.g., Walsh & Rushworth, 1999), and even for studying non-human animals as model cognitive systems that lack linguistic abilities. A similar approach can be used in studying exceptional performers if we assume that they are superior only in certain limited processing abilities; we gain the benefits of working with normal human subjects, but at a cost of being able to draw less powerful inferences.

Second, it is frequently proposed that there are important similarities in the ways that all human abilities develop. For example, at least ten years of preparation are supposed to be required for one to become an expert in *any* field; this corresponds to the time for a child to learn to reliably identify faces, for a student to progress from high school graduate through a doctorate, or for a chess novice to earn the master title (see Ericsson & Lehmann, 1996). However, the fields in which we recognize expertise are those for which only a minority of individuals possess exceptional ability. Again employing the controlled comparison logic, we can compare novices to experts in obscure domains and hope to gain insight into how abilities in common tasks, at which all humans are “experts,” are developed and implemented. Compared to developmental studies, we gain the benefit of adult subjects and discrete, controllable tasks, but we lose the immediate relevance of directly studying language or perception and must settle for indirect inference from chess or mathematics.

Finally, exceptional performance and the limits of human achievement are intrinsically interesting phenomena that are valued by society (“jack of all trades, *master of none*”). As Maslow (1971) reminds us, “if we want to know how fast a human being can run, then it is no use to average out the speed of a ‘good sample’ of the population; it is far better to collect Olympic gold medal winners and see how well they can do.” This is at least as true of mental abilities.

Goals for theories of expertise

What constitutes a good theory of expertise? Four simple criteria suggest themselves. First, the theory should specify the system of processes and structures used by both novices and experts in the domain; the model should ultimately be realizable as a working computer program, and it should fit the constraints of the brain’s basic cognitive architecture.

Second, the theory should specify the critical locus or loci of skill differences within the system. Although a large network of mental and neural components may be required to execute a task as complex as playing chess, each component may be relatively more or less important in determining skill differences.

Third, the theory should specify how the system changes over time to produce improvements in performance. Most research on expertise focuses on differences between experts and novices but pays short shrift to the crucial question of how those differences got there. Recent work suggests that systematic practice, training, study, and coaching are much more important than previously believed (Ericsson, Krampe, &

Tesch-Römer, 1993; Charness, Krampe, & Mayr, 1996), but precise mechanisms of learning would be superior to lists of relevant factors in any theory of expertise.

Finally, the theory should be consistent with previous empirical findings on expertise, and it should directly subsume those parts of previous theories that make sense and have strong support. That is, a new theory need not make a unique, testable prediction about every novice/expert distinction of interest in order for the theory to represent a significant improvement on existing knowledge in the field. Indeed, placing too great a premium on innovation tends to discourage useful theoretical accumulation. If psychology is to make progress by developing mechanism theories that model complex interactions of multiple components of cognition and calibrate parameters to match observed human performance, rather than variable-effect theories that merely catalogue the influence of experimental manipulations on laboratory phenomena, it will have to overcome this problem.

The theory outlined and tested in this thesis does not yet satisfy all of these criteria. In particular, although it directly addresses critical differences between novices and experts, and incorporates key previous findings, it is still too loosely-specified to be implementable, especially as a learning mechanism capable of producing a novice-to-expert shift. This is particularly unfortunate because insight about how expertise develops might in turn help us discover how experts solve problems and how their processing differs from that of novices. These weaknesses should be addressed as the theory is developed further.

Previous research on expertise

Traditionally, research on expertise has not always focused on the theoretical problems discussed above. Instead, it has often been preoccupied with distinctions that are mainly semantic in nature and largely orthogonal to questions about the mechanisms underlying exceptional human performance. For example, in research on intelligence, the issue of heritability versus learning has received a level of attention that is out of proportion to what its resolution could tell us about the mechanisms behind human cognitive abilities, and debates over this issue have clouded thinking on other important aspects of the relevant phenomena (e.g., see Korb, 1994). The same argument can be made regarding the study of sex/gender and ethnic/racial differences in performance; although of intrinsic interest, especially to public policy and the popular imagination, such investigations rarely shed new light on the fundamental mechanisms of behavior.

Other frequently asked questions are whether expertise in specific domains is based on superior perception, memory, or thinking ability, and whether the superiority of experts is specific to the particular task or generalizes to some small or large cluster of related tasks. Again we are dealing in false distinctions and dichotomies when we focus on these issues: where “perception” stops and “memory” begins should be interesting mainly to lexicographers, not psychologists, who need only a useful operational definition that enables empirical research to proceed. Likewise, performance is always produced by a combination of general and specific ability, at least in a factor-analytic sense, and it is unlikely that the mixture is ever 100% of one and 0% of the other. And besides, to answer any of these questions, the best approach will be to elucidate the psychological mechanisms involved before mapping them onto the particular dimensions (perception-memory, general-specific) of interest.

The Standard "Chunking" Theory of Chess Expertise

Research on chess expertise has yielded important discoveries about memory and thinking processes, provides a precisely constrained model environment of a complex cognitive task, and takes advantage of the unique rating system developed by Elo (1961, 1986) that enables precise measurement and comparison of actual skill level between individuals and groups (see **Figure 1**).[1] The canonical theory of chess skill developed from the 1940s through the 1970s by de Groot and by Chase and Simon has had great influence on the field of cognitive science. The success of the standard theory can be attributed to many factors, not least the simplicity, elegance, and counterintuitive but nonetheless satisfying nature of its key features. Charness (1992) documented the widespread inclusion of this work in cognitive psychology textbooks published in the late 1980s, although Vicente and de Groot (1990) and Vicente and Brewer (1993) pointed out how the details of key experiments are frequently misremembered and misreported.[2]

What is appreciated even more poorly than the details of the earlier work, however, is the doubt cast on the standard theory by more recent findings, as well as the important but neglected role of visual mental imagery in chess expertise. The next

---

[1] The Elo system, which was first used in the 1960s, assigns to each player a numerical rating based on his results against other rated players, with the goal of keeping a 200-point difference predictive of a 3–1 winning margin for the higher-rated player (over a long series of games; see Experiment 4 for more discussion of this point). The algorithm for the Elo system is based on the probability models for paired comparisons developed by Thurstone (1927), Mosteller (1951), and Bradley and Terry (1952); for further details see Elo (1986) and Glickman (1995).

[2] In particular, it is often incorrectly said that de Groot (1946) reported the "random chess position" control experiment properly attributed to Chase and Simon (1973a).

sections will briefly summarize the standard theory, its drawbacks, and the case for investigating imagery in chess.

The work of de Groot

Adriaan de Groot (1938, 1946, 1966; de Groot, Gobet, & Jongman, 1996; for an appreciation, see Gobet, in press) was the first to construct a systematic theory of chess skill. His experimental technique was simple: he set before his subjects various chess positions, often taken from his own games (he was a master-level player and frequent member of the Dutch national team), and asked them to think out loud while deciding what move they would play. He wrote down their voluminous comments and analyzed them for various numerical properties, such as depth and trajectory of search, number of moves examined, and so on. The main findings were that compared to the billions of future possibilities contained in a typical game position, players of all levels examined an infinitesimal fraction, about 30–100, and did so in a pattern called "progressive deepening." This was characterized by frequent returns to the top level of the tree and reexamination of already studied variations, rather than long journeys down alternative branches and short backups to prior choice-points, the type of algorithm that would be optimal for a computer program.[3] These results have since been partially replicated (Gobet, 1998a).

Faced with the paradox that players of different skill levels did not appear to behave differently while thinking ahead, de Groot reasoned that stronger players were choosing better moves because they "saw" better moves, not because they were

---

[3] Although de Groot did much of his work before computers even existed, his analysis of tree searching behavior anticipated many later developments in artificial intelligence (see Chabris, 1994b).

calculating more variations. To support this idea he performed his most famous experiment, a simple four-subject test of recall memory for chess positions exposed for only 2–15 seconds each. The results revealed a striking skill effect: the grandmaster was almost perfect, the master and expert were somewhat worse, and the amateur could barely recall half of the piece/square relations. Accordingly, de Groot inferred that skill in chess depended on the ability to rapidly recognize significant patterns (and choose plausible moves directly from them), which was highlighted in the recall test.

The work of Chase and Simon

Chase and Simon (1973a, 1973b; see also Simon & Chase, 1973) replicated de Groot's memory finding with three subjects and extended it by adding a random-position condition in which subjects had to recall positions in which the pieces had been placed on randomly chosen squares. In this case, performance was uniformly low across skill levels, with no advantage for the master.[4] In the normal position condition, subjects were videotaped while they reconstructed the board positions, and the interpiece intervals (the time gaps between placement of successive pieces) were recorded. Analysis showed that relatively long intervals (greater than two seconds) occurred between local clusters of related pieces, but not between individual pieces within such groups. These groups were described as *chunks*, and the master's ability was explained by a short-term memory store holding pointers to several chunk/location pairs which collectively composed the entire chess position.

---

[4] The finding of experts' superior short-term memory for possible configurations but normal short-term memory for impossible or random configurations has been replicated in other domains, such as Go (Reitman, 1976) and electronics (Egan & Schwartz, 1979).

Furthermore, Chase and Simon proposed that a database in the master's long-term memory, perhaps as large as a native English speaker's vocabulary of words, connected the chunks to plausible chess moves, thus completing the explanation of de Groot's observations. Chess expertise, according to the standard theory, is thus based on the acquisition of a large vocabulary of small patterns, each connected via a production-system architecture to specific moves for consideration and analysis; the forward search that follows is important only insofar as it confirms the recommendation of the chunk-recognition mechanism.

Limitations of the Chunking Theory

After de Groot, Chase, and Simon stopped working on chess expertise in the mid-1970s, other researchers began to explore the limitations of their theory (for reviews, see Holding, 1985, 1992; Gobet, 1998b). First, de Groot's own statistics on the search trees generated by his subjects reveal small but consistent advantages for the grandmasters over the experts in such categories as time to choose a move, total moves considered, and mean depth of search. The grandmasters were also significantly faster (over 60%) in examining base moves, or alternatives in the board position under consideration (Holding, 1992). A study by Charness (1981) showed that an expert player searched on average 45% deeper than an amateur player. Holding (1989) found that stronger players improved their judgments as they followed a variation farther from a starting position, but weaker players were equally poor (and inconsistent) at all search depths. By reanalyzing de Groot's original protocols, Reynolds found that de Groot's skilled subjects more frequently used a "homing heuristic" to usefully expand and constrict their search than did his amateur subjects (1991), and that stronger players seemed to attend more to squares attacked by pieces, whereas weaker players focused on squares occupied by pieces (1982).[5] In summary, there appear to be many skill differences in search behavior that de Groot did not discover, probably because his perspective emphasized the restricted nature of the search compared to its absolute limits, a finding that has never been disputed (for example, nobody has argued that grandmasters search thousands rather than dozens of positions).

Questions have been raised about the chunking theory. Both Lories (1987) and Gross (1982, as described by Hartston & Wason, 1985) found that given sufficient exposure time, masters could recall random positions better than could novices.[6] Gobet and Simon (1996a) confirmed a skill difference in the random position condition with a meta-analysis of 13 studies. In a similar vein, Ericsson and Harris (1990) found that repeated practice in memorizing chess positions by a non-playing subject could raise performance on the recall task to expert levels. Additionally, Ginsburg (1995) contends that grandmasters may have larger chunks than the Chase-Simon theory predicts, and that in some cases they may not use chunks at all, recalling complicated positions in a continuous stream of piece-placements without noticeable pauses.

According to the standard theory, since search behavior does not vary with skill level, but chunking behavior does, then the latter must be the source of skill differences in actual chess play. Pattern recognition ability in general is thought to remain relatively

---

[5] Saariluoma (1984) also reanalyzed de Groot's original protocols.

[6] Indeed, as a preliminary step in the procedure of Experiment 5 (see below), we were able to teach 16 masters to perfectly recall four completely random positions, although the process did require many trials and approximately one hour of study (see Chabris, 1994a, for more on memory for random positions).

intact across the lifespan, whereas the capacity for deliberate sequential thought, and working memory in particular, declines with age. Consequently, if skill derives exclusively from pattern recognition, the most skilled players should suffer the least from the aging process, as they will rely more on pattern recognition and less on search relative to less-skilled players. However, Charness and Schultetus (1998) found the opposite in a longitudinal analysis of rating performances by 2500 American chess players: the highest-rated group of players began to lose skill in their 30s, while the second-highest rated group did not decline until their 40s.

But the most important objection to the chunking theory confronts its proposed production-system link between chunk recognition and move selection: small chunks do not appear to have enough independent functional significance to be linked directly to specific moves (see Holding & Reynolds, 1982). In particular, by making a database of visual patterns the exclusive locus of skill differences, the standard theory starkly omits any role for higher-level conceptual thinking. In a 1991 interview I asked grandmaster Patrick Wolff (who later won the U.S. championship twice and spent several years as one of the top 100 players in the world) whether he thinks in terms of general principles or only sequences of moves and countermoves ("variations") during a game:

> Chabris: I find that when I'm playing a game ... that the moves just occur to [me], and I hardly ever think of any overarching recommendations, [as though] there is some kind of process where the general principles are converted into intuition somehow. Do you know of any strong players who ever actually thought in general terms during a game ... as opposed to just seeing variations?
>
> Wolff: *I* think in general terms during the game ... I would be very surprised if there were a single very strong chess player who did *not* think in general terms. It's just that it depends what one means by general terms. I'm not sure that one comes to the chess board and thinks, "OK, this is the *procedure* that I must go through in order to find the right move." I don't think *that* happens.

> Chabris: I mean, more along the lines of "now I want to think about ideas of how to get my rook onto the seventh rank ..."
>
> Wolff: Oh yes, sure, absolutely. And I think it's something that all strong grandmasters have to learn how to do. It's one of the things that separates chess players of a certain class. I'm sure that one of the very things that separates the strongest players from the not-so-strong players is the strength and clarity of that thinking ...
>
> There are several things you have to do in a chess position ... You have to look for your own ideas, and I stress here *ideas.* It's very useful in a position to look for more than just moves, to look for actual ideas: "I want to push this pawn, I want to infiltrate my rook into his position somehow this way, I want to fix his pawns on these squares," and so on and so forth. And if you can get ideas about a position that will lead you to moves.

This exchange points to two clear weaknesses of the approach to chess skill followed over the past 25 years: In pursuit of methodological behaviorism it has neglected the experiences and introspections of the world's top practitioners, and in search of computational elegance it has oversimplified the mechanisms involved to the extent that the ones proposed cannot support the performance levels observed.

It is true that computer simulations of the chunking theory and its descendants (e.g., Simon & Gilmartin, 1973; Gobet & Simon, 1996c) approach the observed performance of chess masters *on the short-term memory tasks the theory was designed to explain.* But the crucial connection between chunks and move selection has hardly been tested at all with human subjects, and no computer program that uses chunk-based pattern recognition to select moves, classify positions, or play a complete game of chess has been successfully developed, let alone demonstrated master strength. For example, Gobet and Jansen (1994) describe their program CHUMP, which narrowed its move selections not to two or three but to 7–20 moves, and actually tended to suggest *more* moves in test positions as its database of chunks grew over time. Mikhail Botvinnik, world champion for 13 years, retired from professional chess to devote himself to building a chess playing program based on human cognition; after 25 years the results were meager (see the critiques of Botvinnik, 1993 by Bronstein, 1993; Berliner, 1993). All of today's successful chess playing computers use virtually no pattern recognition in the move selection process, and no chess software developers argue that a program based on chunking and a human-style limited search mechanism could be successful. It is reasonable, then, to believe that even a program with a database of patterns as large as the repertoire suggested by human chunking experiments would require some additional, crucial elements of human chess cognition to reach human master performance levels.

Visual Mental Imagery and Chess

The first apparent mention of chess as a subject of psychological study is actually a reference to visual mental imagery (Taine, 1875), and one of the best-known works on chess psychology is Binet's (1893, 1894) investigation of the memory abilities of blindfolded chess players (for a review, see Chabris, 1994b). Nevertheless, contemporary research virtually ignores the question of what internal representation chess players are maintaining and manipulating as they "think ahead," or consider what move to play in a given position.

It is clear, however, that chess experts themselves find this to be an important issue. The third world chess champion, José Capablanca (1938) wrote that "besides the modicum of brains, the game [demands] an additional quality—that of a rich imagination." Was Capablanca referring to visual imagery per se, or to creativity? Probably the former, as the American grandmaster and psychologist Reuben Fine (1965) asserted that "there is no question that blindfold chess depends on the capacity to visualize the board with full clarity.” More specifically, the British master Abrahams (1952) argued that “the most important mental activity in Chess is vision, by which is meant the unforced intuition of possibilities by the mind’s eye,” and world championship challenger David Bronstein wrote that “visual images in solving processes play practically the leading role ... in high-level chess thinking the image component, the complete structural vision of the properties of a chess position in movement, occupies a leading place” (Bronstein & Smolyan, 1982). But what is the nature of the image chess players visualize? Is it a veridical, three-dimensional image of a particular chess board and set of pieces (perhaps the ones currently being played on, or the ones the player most often uses)? Taine (1875), based on the report of one amateur player, believed that it was, and characterized the type of imagery used in chess as an “internal mirror” that reflected the precise state of the thing(s) being imagined.

But further examination of the chess literature suggests a different conclusion. Elaborating on his previous remark, Fine (1965) said that “the visualization that takes place must emphasize the chess essentials and must eliminate accidental factors, such as size or different colors [of the board and pieces].” Commenting on the qualities of the chess player’s internal image, Jacques Mieses (1940) wrote that “it is not a planimetrical or stereometrical picture that appears before the mind’s eye of the chess player ... His task lies rather in mentally picturing the constantly changing formations of the pieces ... This process is indeed more closely allied to the department of ‘topographic sense’.” And one of Binet’s subjects, a young master named Goetz, reported that he was “aware only of the significance of a piece and its course ... to the inner eye, a bishop is not a uniquely shaped piece, but, rather, an oblique force” (Binet, 1893).

The plethora of comments on the nature and importance of visualization in the chess literature, of which the above are merely a sample, contrasts sharply with the near-total absence of mention of pattern recognition from the same sources (this changed somewhat in the 1980s, when chess players had become aware of some of the results of Chase and Simon). Of course, introspective reports should not be taken as conclusive evidence regarding mental processes, but neither should they be ignored entirely. Instead, they can serve a heuristic function, suggesting avenues of exploration and elements to include in theories, and they set "boundary conditions"—theories of cognition that cannot account for introspective experience cannot be considered complete (see Chabris, 1999, for further discussion). In the case of chess, the reports of chess masters suggest that mental imagery may play an important role in expert thinking and may be a locus of differences between experts and novices. The next section will propose a theory adding visualization processes and conceptual knowledge representation to the chunking theory to develop a more accurate model of expertise in chess.

The Mental Cartoons Hypothesis

Given that the standard theory of chess skill is unable to account for some aspects of expert performance and neglects the role of mental imagery, a new theoretical framework may be warranted. Inspired by the comments of chess experts on their own experiences, I propose the "mental cartoons" hypothesis: *Expertise in visual-spatial domains such as chess is based on the development of cartoon-like representations of the domain's important properties, as contrasted with photograph-like representations of the domain's constituent elements.*

The notion of mental cartoons, which will be elaborated and tested in this thesis, is consistent with research in face perception that proposes that the mind represents familiar faces as caricatures or "superportraits" that emphasize their distinctive or unique features (Rhodes, 1996). It is also reminiscent of research on physics, where experts appear to solve problems by thinking in terms of the abstract concepts (point masses, friction) involved in a given situation rather than the objects (block, inclined plane) present in the display[7] (for a review, see Anzai, 1991). The hypothesis in its most extreme form implies that if we study experts in any visual-spatial activity, we will discover that they possess an underlying cartoon representation for the domain, or at least that characterizing and simulating such a representation will help us successfully understand and model the behavior of experts in the domain.

---

[7] According to Anzai (1991, p. 88), diagrams drawn by physics experts "tend to be principle-oriented abstractions of physical objects and their relations that are quite different from realistic drawings." That is, they draw cartoons, and to the extent that drawings produced in problem solving correspond to the internal images used by the problem-solver, they appear to represent problems in cartoon form. For information on possible cartoon-like sketching practices of architects, see Robbins (1994).

Two distinct aspects of the cartoon metaphor are central to this view of expertise. First, cartoons systematically distort spatial relationships; second, they highlight important information and obscure unimportant information. **Figure 2** illustrates the main components of the proposed theory, as applied to chess. Perceptual input initially creates a largely veridical representation of the environment that preserves detailed, high-resolution, three-dimensional information. This is automatically recoded into a two-dimensional "semi-depictive" diagrammatic representation, the mental cartoon. This structure preserves only the information needed to perform within the rules and constraints of the domain, so it may distort spatial relations if necessary, and it is the arena in which visualization of future possibilities takes place (cf. the "visual-spatial sketchpad" of working-memory theory; Baddeley, 1986). It in turn feeds forward to separate stores of domain-specific pattern and conceptual knowledge, which interact with one another and feed backward to affect the contents of the cartoon. This top-down influence causes the cartoon to emphasize and de-emphasizes different areas, items, or configurations as problem solving progresses.

The mental cartoons hypothesis may be seen as part of a solution at the computational theory level (Marr, 1982) to the information processing problem of making high-quality decisions in "combinatorially explosive" situations. These are tasks in which a set of elements may be combined and recombined in so many different ways that it would take an organism almost forever to examine all of the possibilities. For practical purposes, the problem is underconstrained, just like the problems of color vision or object identification, which cannot be solved accurately 100% of the time by any system, biological or artificial. A mental cartoon's functions are to strip away irrelevant detail, magnify relevant distinctions, and thereby accelerate processing within the domain, perhaps to a point that is in some sense optimal for the human mind, given the time and space restrictions imposed by the brain's physical and computational architecture (cf. Ericsson & Lehmann, 1996). This involves making assumptions about how to organize the domain's elements that will prove useful in the majority of situations, but which may fail in pathological cases, just as the visual system's assumption that L-junctions in an image denote corners in a physical object causes an erroneous perception of the Penrose triangle.

Principles of Mental Cartoon Theory

Mental cartoons have six key properties, which are explained below using chess as the primary example. One general principle of mental cartoons, however, is that each domain's cartoon will contain and emphasize those properties that are necessary for that task, so not every aspect of the chess cartoon will have an analogue in the mental cartoon representation used for every other domain. Also, a cartoon corresponds to a particular class of problems or computational task, not necessarily to the entire set of activities performed by a professional in a domain, which may be quite different and rely on diverse processing mechanisms. (For example, a graphic designer may use one type of mental cartoon when laying out the elements of a magazine advertisement and a completely different one when drawing the letters of a new typeface.)

Some of the properties and distinctions described below will be examined further in the experiments presented in this thesis. Several other relevant experiments will be proposed for future consideration.

Mental cartoons are semi-depictive

Mental cartoons are spatial representations that lack some, but not all, properties of standard mental images. By "standard mental images" I mean the type of image studied by the canonical experiments on visual mental imagery, such as rotation, scanning, and generation paradigms (for reviews, see Kosslyn, 1980, 1994). Considerable research has been devoted to showing that, like pictures, these images share many properties with the actual physical stimuli they represent and/or the visual percepts of those stimuli. For example, when an object is rotated through space, it is constrained to pass through the intermediate angles between the start and end points; the finding that mental images of objects obey the same constraint is powerful evidence that imagery is a depictive, or picture-like, representation system. By contrast, a non-depictive, or propositional, system need not obey the rotation-time constraint, since an object's pre-rotation configuration can be transformed mathematically into its post-rotation configuration directly, with no time spent on intermediate angular configurations. Mental cartoons are a third code: they are depictive representations in some ways, but they are distorted depictions of physical stimuli, in that they do not obey all of the usual imagery constraints. There is ample precedent for the brain's use of codes that do not correspond perfectly to our perceptual experiences or to the conventional systems we use to communicate. Spatial frequency, population coding, coarse coding, and end-stopped visual neurons are such codes; the mental cartoon (or more generally, the caricature that exaggerates distinctive or important features) may well be another.

In chess, the cartoon should normally have the form of a two-dimensional symbolic diagram, not a three-dimensional picture. As Fine (1965) pointed out, correct behavior in the chess domain is defined purely by the identities of the pieces and squares, not their peculiar shapes, colors, distance from the viewer, shadows, etc. Thus an efficient encoding of the board position will strip away those factors, resulting in a basic representation similar to chess diagrams found in magazines and books.[8] (Computational comparisons, using an objective measure of visual similarity, may show that pairs of two-dimensional symbolic diagrams of chess positions are less similar than pairs of photographic images of the positions.) One preliminary test of this idea could involve asking chess masters to close their eyes, imagine a particular sequence of moves, and then to answer a question about the resulting position; those answering correctly would then be asked whether they were maintaining an image of a three-dimensional chess set and board or of a two-dimensional diagram. Or, after the imagination exercise, the masters could be given a priming test to see if two-dimensional symbolic or three-dimensional photographic depictions of chess pieces were easier to name (as compared to a baseline before thinking ahead). Kosslyn, Anderson, and Chiu (in preparation, as described by Kosslyn, 1994, p. 288) found similar priming for objects that were previously visualized by participants, but not for objects that were mentioned in sentences generated by participants. Finally, we could give chess masters and novices a recognition memory test with photographs of chess positions taken from various angles and distances and show that the masters were able to generalize from previously seen

---

[8] It is of course possible that the internal representation may resemble printed diagrams only because chess masters have experience in perceiving and imagining such diagrams from the literature. The direction of causation here might be tested with a developmental approach, using children learning chess for the first time and who had not been exposed to printed chess literature.

views to new ones more accurately than the novices, as would be predicted if they were using a viewpoint-independent internal code.

The cartoon is normally generated from another representation, either top-down or bottom-up; in chess it would ordinarily be composed from lower-level visual representations created by input from looking at a chess position. The cartoon feeds back to this "perceptual image" and can be refreshed from it as long as it remains present; however, low-level input is not required, as the phenomenon of blindfold chess shows.

For a demonstration study in 1998 grandmaster Sergey Kudrin reconstructed positions with a standard set and board after five seconds of viewing them as printed chess diagrams. Afterwards, I asked him how he remembered where the pieces were located. "I visually remembered the diagram, but [the pieces] were on this diagram, which for me is almost the same as the board, but in my mind they stayed as the pieces of this diagram," he explained. I asked if he had any problems in translating between his memory of the diagram and the three-dimensional board and pieces. "It seemed natural, although maybe I lost it somewhere because I didn't get this [particular] position exactly. Although at the moment when I stopped looking at [the diagram], I was sure I would remember everything, but I didn't remember, and I knew I was doing something wrong," he replied. That is, Kudrin formed some sort of visual image of the chess position with the same characteristics as the two-dimensional diagram he studied, and when the stimulus was removed he had a clear image. But as he began to construct the position on the board, the image faded, and he could not complete the task as accurately as he initially expected (cf. the reports of subjects in the "whole report" visual storage paradigm of Sperling, 1960). These introspections, to the extent that they are accurate, contradict the notion that performance in the recall task reflects purely the organization of a subject's memory for the position. At a minimum, the chunk level of representation interacts with a full-scale visual image of the position, but a more speculative possibility is that the bursts and pauses in recall by very skilled subjects reflect focusing and shifting attention at various local areas of that rapidly fading image, and are therefore more artifactual than diagnostic with respect to memory organization. Before such a conclusion can be sustained, however, further research is needed.

When the cartoon is modified, as when a chess player imagines a certain move being made, the perceptual image is not necessarily modified as well (see discussion below). The progressive deepening search pattern found by de Groot (1946) suggests that the cartoon fades quickly and must be refreshed frequently by restarting analysis at the current board position, normally available continuously in visual input.[9] Accordingly, it may be of interest to obtain protocols from blindfold play and ask whether they contain more backtracking to the root position than do normal protocols.

Another distortion is the incomplete preservation of spatial relations. In chess, the physical distance between two squares is not as important as the number of intervening squares; for example, the square at the upper-right corner of the board is just as "far" from the square at the lower-left corner as it is from the square at the lower-right corner, since they are each six squares away, although the former lies on a diagonal path approximately 1.4 times longer in physical distance than the vertical path to the latter. Therefore, although direction will be represented in the chess cartoon, distance will be

---

[9] Former world championship candidate Artur Yusupov (1998, pp. 154–155) explained one of his errors in this way: "The reason for such a curious error must lie in my indistinct image of the chessboard while I was analysing; a vital detail of the position was simply excluded from my internal field of vision."

measured using a less-than-purely depictive metric. The consequences of this property will be addressed later.

Mental cartoons are domain-specific

These distortions are not general changes in mental imagery abilities, but are specific to the domain of expertise. Thus, a chess player will gradually develop a cartoon representation of the chess domain that will help him to play better and better chess, but it will have little effect on his ability for playing backgammon, cooking, or picking stocks, and it will not prevent him from accurately comparing diagonal and vertical distances. However, insofar as some aspects of the domain appear in other contexts or skills, transfer may occur. For example, a chess expert may have less difficulty than a chess novice when asked to visualize a board position in checkers, since both games rely on the same geometrical structure (the 8x8 grid of alternating light and dark squares) for their basic representation.

Indeed, the second world chess champion Emanuel Lasker (1932) urged chess players to train themselves in visualizing a completely *empty* board:

> It is of importance that the student of Chess should know the board very accurately; he should be able to visualise each square in its individual position as well as in its relations to its neighbouring squares ... The student should endeavour to acquire the habit of designating the squares and of visualising their position. There are many Chess-players who fail merely from their incapacity to master this geometrical task, not suspecting its value. (pp. 3–4)

It should be noted that the domain-specificity of mental cartoons does not preclude positive transfer to other tasks from experience in chess. Factors of motivation, concentration, attention, and the like might be affected by exposure to chess (as considerable anecdotal evidence suggests), even though they are not the cognitive loci of skill differences.

Mental cartoons facilitate important operations in the domain

The distortions that distinguish mental cartoons from mental images are not random perturbations, but systematic changes that facilitate important operations in the domain. These operations may involve one or more visual manipulations (rotation, generation, comparison, etc.), depending on the nature of the task. The mental cartoons of someone with expertise in information graphics, for example, might facilitate some of the processes described by Simkin and Hastie (1987) in their task decompositions of bar- and pie-graph understanding. In the case of chess, the crucial transformation is the rapid and accurate imaginary execution of a legal chess move. Recall that although de Groot emphasized the similarity in search behavior between novices and masters, his own data show that grandmasters tend to examine moves at a faster rate than masters, and subsequent studies suggest that experts can search significantly deeper than novices in a given amount of time (Charness, 1981).

In chess, the actual “length” of a move, or the number of squares a piece crosses to get from one square to another, is not a functionally useful measure—if a move is legal, it costs the same as any other move and is just as “afforded” by the position no matter how long it is. Thus, we might expect that the chess cartoon will remove, or at least attenuate significantly, the distance-time constraint typically found in image scanning (e.g., Kosslyn, Ball, & Reiser, 1978): the time per square traversed to mentally execute a legal move should be lower for chess experts than for chess novices, with no difference in the time to encode the stimulus or make the response. Since this effect is not predicted by any other theory of chess skill, it will be of special interest to explore it further (see Experiment 1).

Mental cartoons interact with stored knowledge

A mental cartoon is not an unstructured montage. It has a basic dimensionality and boundaries as well as a hierarchical superstructure that groups its elements into increasingly abstract sets, culminating in units that can represent families of configurations for the entire cartoon (an idea already sketched out by Bergson, 1902, in a critique of Binet's imagery work). This is a generalization of the chunking aspect of the standard theory; in chess the basic cartoon structure would be an 8 x 8 grid and the hierarchy could be formed by chunks, "chunks of chunks," and so on up to what might be considered fuzzy, overlapping schemas that represent large categories of board positions. This kind of chunking has more in common with the way the term is used in artificial intelligence (e.g., Campbell, 1988) than in psychology. Thus, mental cartoon theory is consistent with recent ideas about "long-term working memory" and retrieval structures (Ericsson & Kintsch, 1995), it predicts the same results that chunking theory predicts, and it predicts results of tasks that chunking theory cannot address.

This hierarchy could be implemented by an interactive activation model like the one proposed by McClelland and Rumelhart (1981) to account for word recognition phenomena. Laterally inhibitory representations at high levels (position categories) can feed activation down to low levels (chunks and pieces) as well as receive activation up from them. Thus, we might observe a "chess position superiority effect" analogous to the word superiority effect: chunks or even individual pieces embedded in complete positions would be easier to identify than chunks or pieces presented without context—but only when the complete position is normal, not random—just as letters are easier to identify within words but not pseudowords because of a top-down contribution to the activation of their representations (Rumelhart & McClelland, 1982). Regardless of how the non-chunk aspect is implemented, the addition of conceptual knowledge that cannot be described purely in terms of first-order visual patterns is a key aspect of this theory. This knowledge, in addition to being directly accessible during the thinking ahead process, projects backwards to amplify important parts of the cartoon data structure.

An important aspect of mental cartoon structure is that some components of the cartoon will be more or less likely to change over time (during a single problem-solving episode or series of episodes, such as a chess game). In chess, some pieces move "faster," or to more distant locations in fewer moves, than others. Those that remain fixed for long periods, such as formations of pawns or the locations of the kings (which can move only one square at a time), will be more likely to define higher-order schemas and suggest higher-order, longer-running plans of action. A consequent and untested prediction is that thinking ahead in chess over sequences of moves including exchanges of pawns should be more difficult and error-prone than imagining equally lengthy sequences containing only piece (non-pawn) moves. Grandmaster Jonathan Tisdall (1997) in fact recommends several blindfold chess visualization exercises in which the difficulty is keeping track of changing pawn locations. In the standard theory, chunks are linked directly to suggested moves via a production system, with no role for intervening goals, strategies, plans, or conceptual information of any kind (see Experiments 2–4).

Mental cartoons are dynamic

A mental cartoon is not a static representation that requires intervention by outside processes to change. Rather, it has properties of both a structure and a process, and its contents can change automatically according to specific movement patterns of its elements. For example, once a sequence of moves has been imagined in a chess position, the sequence may replay itself over and over again like film loop as the player analyzes its consequences and verifies its accuracy. Operation of the cartoon itself as well as external calculation procedures each have roles in determining how its elements will be combined and recombined.

Recall that Binet's subject Goetz described a bishop as not an object but "an oblique force." Likewise, grandmaster Nikolai Krogius (1976) describes the chess image as containing "lines of force," and Steiner (1972, p. 66) says

> The great chess player does not see squares and pieces as discrete units, or even abstract counters. He internalizes a very special sense of "fields of force," of regions characterized, differentiated by the fact that certain events can or cannot take place in them. What matters, thus, is not the particular square, or even piece, but a cluster of potential actions, a space of and for evolving events.

Grandmaster Paul Keres (1991, p. 69) described a key moment in one of his own games in similar terms: "During the game I had the feeling that something must be happening around [the square] d5 ..." Note the congruence between these observations and the results of Reynolds (1982) regarding the importance of empty space to strong players.

But why should the chess player's introspections reveal forces and actions, which would seem too abstract to be pictured mentally, rather than pieces and squares? A theory of chess expertise should be able to answer the question of what the mind's eye is looking at in such cases. Perhaps the dynamic activity of the pieces in the cartoon (the "unforced intuition of possibilities" cited by Abrahams), when fed back into earlier visual areas, results in a kind of apparent motion on the perceptual image of the chess board and pieces. As a piece jumps back and forth between two squares in the cartoon, it will appear to move in the perceptual image, just as a spot of light alternating regularly between two separated locations in the world will appear to move through the intermediate positions (in our percept). The accumulation of these movement patterns over a few seconds' time might appear to be "lines of force," just as true apparent motion leaves more of a fuzzy trail than a series of distinct frames. There are other possibilities; for example, the cartoon structure may include not only visual but also motor imagery, or a sort of "gesture."[10]

Mental cartoons can impair as well as improve performance

Some tasks involving elements of the cartoon domain in unusual ways may actually be performed relatively worse with the cartoon representation than without it (i.e., by an expert than by a novice). The results of Bachmann & Oit (1992) discussed below, as well as the recent findings that bartenders and waiters may be worse than laymen at Piaget's water-level task (Hecht & Proffitt, 1995; cf. Vasta, Rosenberg, Knott, & Gaze, 1997) and that golfers misestimate the weights of golf balls (Ellis & Lederman, 1998), may be consistent with this idea. At any rate, unless mechanisms supporting

---

[10] The artist Marcel Duchamp, who was also a strong international chess player, told an interviewer: "In chess there are some extremely beautiful things in the domain of movement, but not in the visual domain. It's the imagining of the movement or of the gesture that makes the beauty, in this case" (Cabanne, 1971, pp. 18–19). On the distinction between conceptual knowledge and visualization in chess thinking, he said, "I believe that every chess player experiences a mixture of two aesthetic pleasures: first the abstract image akin to the poetic idea of writing; second the sensuous pleasure of the ideographic execution of that image on the chess board" (Duchamp, 1952, as quoted by Kremer, 1989, p. 39).

exceptional performance are always applied with perfect specificity, they will sometimes be used in situations for which they were not designed, and errors may result.

Additionally, since mental cartoons represent information neither pictorially nor propositionally, nor by a dual code combining the two, but rather by a “third code” of sorts, individuals relying on mental cartoons for expertise may have difficulty communicating their specialized knowledge to others. For example, a plan of action in a domain whose internal representations are mainly propositional can be written down as a series of instructions; in a purely spatial domain it can be shown as a sequence of pictures or a film; in a cartoon-based domain, however, it cannot be conveniently expressed either verbally or pictorially. This difficulty of transferring cartoon-encoded information between individuals might help explain why group problem-solving and brainstorming exercises are more successful in some domains than others.

Overview of Experiments 1–4

In this section I have outlined a broad and somewhat speculative theory of expertise in chess and other domains. With respect to chess, it goes beyond the classical chunking theory by giving a key role to visualization, in the form of the mental cartoon itself, and by distinguishing between knowledge represented as local chunks and patterns and knowledge represented by higher-order, conceptual structures, both of which support processing in the cartoon by bringing important information to attention. I have mentioned many results that indirectly support this theory, and I have suggested several experiments to test its predictions more directly. Experiments 1–4 do this as well.

The first experiment concerns the role of visualization in chess expertise and whether there are skill differences in domain-specific image manipulation that are associated with essential chess functions. This addresses the semi-depictive nature of the mental cartoon. The next two experiments provide evidence that chess masters represent chess knowledge, specifically about famous chess positions, in a different form from a list or map of chunks. The fourth experiment tests a key prediction of the chunking theory, and finds that contrary to that theory's emphasis on rapid recognition as the primary source of expert ability, the decision quality of chess grandmasters improves significantly with additional thinking time.

Experiment 1: Image Scanning

Some investigators (e.g., Eglen, 1993) have claimed that verbal processing is important in chess skill, and one could argue that mental cartoons are not really a third code but a combination of visual and verbal codes (cf. Anderson, 1978), each used when optimal for the task at hand. However, concurrent simple articulatory processing does not significantly impair chess players' recall memory or quality of move selection. By contrast, secondary tasks designed to interfere with the "visual-spatial sketchpad" or "central executive" components of the working memory system both disrupt chess memory and decision-making (Robbins et al., 1996; see also Saariluoma, 1992). These results suggest that mental cartoons in chess are based more on visual and abstract representations than on verbal ones.

The idea that mental cartoons accelerate certain processing operations in the domain has been touched on by several experiments. First, Bachmann and Oit (1992) asked novices and experts to imagine a chess piece following a sequence of one-square up/down/left/right movements on a chess board. When the given piece could not legally move in that fashion (e.g., a bishop, which only moves diagonally), expert subjects were less accurate at reporting the final location of the piece than they were with a piece that could make the moves (e.g., a rook); novices showed no effect. Overall, the experts performed better than the novices. These results are consistent with the claim that mental cartoons in chess facilitate the transformations involved in making legal moves, but not illegal ones.

Dallenbach (1917) reported a detailed introspection, in the classical style, of his mental experiences while playing blindfold chess. Regarding the operation of executing chess moves in a mental image, he made several observations consistent with the mental cartoon idea:

> When a piece or pawn is moved from one square to another, the visual images involved are of two kinds. Either the visual image of the piece moves across the face of the board, or the piece occupies the new square without having been visualized in passing. The second of these experiences is the more common. When movement across the face of the board is seen—and this is usually the case in moves involving long jumps—the piece seems to be moved by an invisible force ... The force which causes the imaginal pieces to move about the board seems to be vested in them; they are not moved by any external power, such as an imagined hand. (p. 222)

The most interesting empirical study on this topic was done by Milojkovic (1982). He showed subjects a simple chess position containing three pieces in a triangular formation and, after removing the position, asked them to mentally execute a given move. When they had done this they were to press a button, at which point a new position would appear and they had to decide whether it correctly showed the result of the move they were just asked to imagine. All of the moves were captures, and they were cued by colored cards whose meanings (e.g., blue = "queen takes rook") had been memorized previously. The distance separating the pieces varied between 1 and 5 intervening squares. The key result was that the lone master subject displayed an extremely shallow per-square time to imagine the moves—5 millisecond slope of best-fitting line—especially compared to the novice subjects, whose average slope was closer to 30 milliseconds per square. The magnitude of this effect is comparable to those taken in visual search tasks to confirm parallel processing instead of serial search, and suggests that the scanning operation may be accelerated for chess masters when it corresponds to the path of a legal chess move. If true, this would provide strong support for a key prediction of the mental cartoons hypothesis that is not addressed at all by the standard theory, which says nothing regarding the speed of mentally executing moves.

Milojkovic's result, while provocative, is somewhat problematic. It is based on a comparison of only one chess master to a group of novices, and the number of observations per participant was limited. More importantly, his master may have been able to use a verbal encoding to enumerate which captures were possible and which were not during the study phase of each trial, so that no visualization would actually be required during the move-execution phase. The goal of Experiment 1 was thus to use a superior paradigm to test whether expertise in chess is associated with superior image-scanning ability that is specific to the visualization of chess moves.

Finke and Pinker (1982, 1983; see also Pinker, Choate, & Finke, 1984) proposed that the typical image scanning behavior pattern would occur spontaneously if it was the only way to solve a problem. They presented displays with several arbitrarily located dots for several seconds; after the dots disappeared and a retention interval passed, an arrow would appear briefly, and participants were to decide if the arrow would be pointing at any of the dots if they were still present. Response times increased with the distance of the nearest dot (if any) along the path of the arrow, indicating scanning. Importantly, since participants had no way to predict where the arrow would appear, they could not use the study interval to code the stimulus in such a way as to permit instant responding when the probe appeared.

We adapted this paradigm to chess by creating chess diagrams with four black pieces arbitrarily distributed on the board. As in the Finke-Pinker setup, participants were to study these displays and then, after a retention interval, decide whether a single displayed white piece could capture any of the black pieces if they were still present. We predicted that chess masters would be able to scan their retained images faster than chess novices, but that the two groups would not differ in how fast they scanned in the original Finke-Pinker conditions.

Method

Participants

Six chess masters and 18 chess novices volunteered to participate and were each paid \$10.00 for a testing session of approximately one hour.[11] The masters were three Harvard University undergraduate students and two faculty members, plus grandmaster Patrick Wolff. Their January 1999 U.S. Chess Federation (USCF) ratings ranged from 2199 to 2650 (mean 2352, standard deviation 163), and their ages ranged from 18 to 32 (mean 25.3, standard deviation 6.6). The novices were Harvard University undergraduate students who had no ratings and reported never having played in organized chess tournaments; their ages ranged from 18 to 22 (mean 20.1, standard deviation 1.3). All participants were male; five-sixths of the members of each group (5 in the master group, 15 in the novice group) were right-handed as measured by the Edinburgh Handedness Inventory (Oldfield, 1971).

---

[11] The data from three other participants was excluded because two committed more than 40% errors on the Yes trials of one of the two tasks, and we could not verify the rating of one who claimed to be a master.

Materials and Design

Black-and-white stimuli were designed for two analogous versions of the task: The “Chess” task and the “Dot” task. **Figure 3** shows the stimuli and event sequence for a sample trial in each task.

The chess task stimuli were approximately square diagrams of a chess board (7.3 cm wide x 7.6 cm high), similar to those found in newspaper columns, magazines, and books about chess, created with the Linotype Game Pi ChessDraughts typeface (Adobe Systems, Mountain View, CA). 80 study diagrams were prepared by distributing the symbols for four black chess pieces (queen, rook, bishop, knight) on the board in a pseudo-random fashion. For each of these diagrams, a corresponding probe diagram was prepared, showing only one white bishop with an arrow (7.1 mm long) pointing out of it along a diagonal trajectory (thus corresponding to a legal chess move, as bishops move only diagonally). Half of these probes pointed at a location occupied by a black piece in the corresponding study diagram (“Yes” trials), and half pointed at least 45 degrees away from the nearest direction that did point at a black piece (“No” trials).[12] Of the 40 Yes trials, there were eight each with one, two, three, four, and five intervening squares between the white piece probe and the black piece it would be pointing at. The 80 probe arrows pointed equally often in the four diagonal directions (45, 135, 225, and 315 degrees), and this was counterbalanced across Yes and No trials; angle could not be counterbalanced within each distance because there were not enough trials. The arrows pointed only diagonally because pilot testing suggested that scanning behavior was more

---

[12] This makes the task easier than if the arrow probes could be “near misses.” We were following the original design of Finke and Pinker in this respect; it did not prevent them from finding scanning behavior.

consistent in the dot task with diagonal arrows.[13] The five distances corresponded to 1.3 cm per square, or 1.3, 2.6, 3.9, 5.2, and 6.5 cm for 1 through 5 squares. Finally, one diagram of an empty chessboard was prepared for display during the retention interval between study and probe stimuli.[14]

The stimuli for the Dot task corresponded one-to-one with the stimuli for the Chess task. We erased the chessboard background from each diagram, replaced the black pieces in the study diagrams with black discs approximately 5 mm in diameter, and erased the white piece from the probe diagrams (leaving the arrow intact).

Thus, in each task there were two independent variables of interest: correct response (Yes or No) and, for the Yes trials only, distance (1–5 squares).

Procedure

Participants were tested individually in a single session lasting approximately one hour. After giving informed consent and completing various forms, they completed the Dot task and then the Chess task. This order was chosen because if participants completed the Chess task first, they might have tried to adopt the strategy in the Dot task of coding the stimuli as though they were chess pieces on a chess board. Psyscope 1.2 software (Carnegie Mellon University, Pittsburgh, PA) running on an Apple Power Macintosh computer with a 15-inch display was used to present stimuli and record

---

[13] Also, Church and Church (1983) showed chess displays that did not contain a checkerboard background to a strong amateur participant and found a steeper slope for imagining diagonal than horizontal and vertical moves.

[14] In half of the study diagrams, each black piece was "protecting" (according to the rules of chess) at least one other black piece, while in the other half, no black piece was "protecting" any other black piece. However, the data were collapsed across this factor because there were too few trials of each distance to allow meaningful comparisons.

responses and latencies. Participants sat at a viewing distance of approximately 50 cm from the computer screen; thus, the diagram or box stimulus subtended a visual angle of approximately 8.3 degrees horizontally by 8.6 degrees vertically. Each task began with instructions, which demonstrated the sequence of events in a trial, and ten practice trials that familiarized participants with the stimuli and task, followed by 80 experimental trials. Participants were allowed to pause after 40 trials, and were asked to rest for a few minutes between the two tasks. After completing both tasks they answered several questions about their experiences during testing.

The procedure was identical for the two tasks. The 80 trials were presented in a pseudo-random order such that no correct response (Yes or No), distance (1–5, for Yes trials only), or arrow angle (45, 135, 225, or 315 degrees) could appear more than three times in a row. Each trial consisted of the following events: (1) a large asterisk, which served as a warning signal and a fixation point, appeared in the center of the screen for 1000 ms; (2) a study stimulus appeared in the center of the screen for 5000 ms; (3) a blank diagram (Chess) or empty box (Dot) appeared in the center of the screen for 2000 ms; (4) a probe stimulus appeared in the center of the screen. Participants responded by using their right index finger to indicate “yes” by pressing the B key or their right middle finger to indicate “no” by pressing the adjacent N key on the computer keyboard. Participants were instructed to respond as quickly and accurately as they could. As soon as a key was pressed the trial was over and the next one began. Note that the sequence of events 2–4 described above appeared to participants as a chess diagram (or box) that first contained four black pieces (or dots), then contained nothing, and then contained a single white piece with an arrow (or just an arrow); there was no “flicker” or inter-event blank.

Results and Discussion

We analyzed the data from each task in the same way. First, we grouped each participant's trials according to correct response, Yes or No. We further sorted the Yes trials according to the distance between the arrow and the location of the dot or chess piece at which it pointed. Within each of the resulting six cells (Yes-1, Yes-2, ..., No), we calculated the percentage of errors made, and discarded the response times for these incorrect trials. We trimmed the remaining response time (RT) data points in each cell by iteratively selecting the longest response time and removing it if it was more than three standard deviations above the mean of the remaining RTs in the cell. This procedure makes RT distributions more normal and excludes trials on which participants probably either lost concentration or did not follow instructions; in this case it eliminated 2.79% of all collected data points. Finally, we calculated the mean RT for each participant for each of these cells.

For each participant, we regressed the Yes RTs onto the scanning distances (1–5 squares) and calculated the slope and intercept of the resulting "scanning function," or best-fitting line. The slope measures the speed or efficiency of the image scanning process in milliseconds per square, whereas the intercept specifies the time consumed for all cognitive operations *except* scanning, and measures the efficiency of encoding the arrow probe, preparing to scan, and producing a "yes" response. Thus, the slope and intercept are independent behavioral measures that need not covary. For each task and group we averaged these parameters across participants, and also calculated the correlation *r* between the group average RT for each distance and the distance itself. This assesses the linearity of the group scanning function.

The predictions made by the mental cartoon theory were borne out, as shown in **Figure 4**. First, in the Chess task, the masters had a scanning function slope less than one-fourth as great as that of the novices (43 versus 180 ms/square, $r = .93$ and .99), $t(22) = 2.94$, $p = .008$, $d = 1.42$. However, the masters had a non-significantly higher scanning function intercept than the novices (1085 versus 918 ms), $t(22) = 1.32$, $p = .199$, $d = 0.62$. Taken together, these two results suggest that masters process mental images of chess configurations more efficiently and that this advantage is specific to the mental imagery process involved in visualizing legal chess moves, rather than general to processes of encoding and representing chess configurations (and responding). This conclusion is reinforced by the lack of a speed advantage for masters over novices in responding to No trials, on which there are no legal chess moves to be visualized (1141 versus 1021 ms), $t(22) = 1.07$, $p = .298$, $d = 0.52$. Moreover, although the mean Yes-trial slope for the masters was positive (43 ms/square), it was not significantly different from 0, $t(5) = 1.52$, $p = .189$, $d = 0.62$; indeed, for three of the six masters (and none of the 18 novices) the slope was less than 10 ms/square.

Second, on the Dot task, there were no significant differences between masters and novices in either scanning function slope (47 versus 58 ms/square, $r = .53$ and .83), $t(22) = 0.25$, $p = .803$, $d = 0.12$; scanning function intercept (1406 versus 1113 ms), $t(22) = 1.51$, $p = .145$, $d = 0.73$; or response time on No trials (883 versus 925 ms), $t(22) = 0.48$, $p = .633$, $d = 0.23$.

We ruled out the possibility of speed-accuracy tradeoffs in the data: There were too few errors for meaningful comparison in the Chess task, where the overall error rate was 2.08%, and there were no significant differences in error rates between masters and novices on either Yes or No trials in the Dot task, $p > .05$ in each case.

We also compared the scanning functions on the two tasks separately for each group. The slopes for the novices were higher in the Chess than in the Dot task (180 versus 58 ms/square), $t(17) = 3.788$, $p = .002$, $d = 1.84$; this difference was not significant for the masters (43 versus 47 ms/square), $t(5) = 0.110$, $p = .917$, $d = .10$. The interaction between task and group was significant for the slopes, $t(22) = 2.113$, $p = .047$, $d = .90$. For the intercepts, the same interaction, as well as the differences between tasks for each group, were not significant, $p > .05$ in each case. These results bolster the conclusion that masters scan chess images differently from novices. It should be noted that the general (though non-significant) advantage for the novices over the masters in both scanning function intercept and No response time (except in the Dot task) could reflect the age difference between the groups: the masters were significantly older than the novices (25.3 versus 20.1 years), $t(22) = 3.35$, $p = .003$, $d = 2.03$, and might thus be expected to respond more slowly on average.

To explore this difference further, we computed each subject's score on the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1972, 1973), which was completed before the task began. The VVIQ asks subjects to visualize different scenes, such as "some relative or friend whom you frequently see" or "a rising sun," and rate how vivid various aspects of each image seem on a 1–5 scale. Two novice subjects did not complete the questionnaire; for the remaining 22 subjects chess masters experienced *less* vivid mental images, a difference of 10 points out of a total of 80 possible (63 for novices versus 53 for masters), $t(20) = 2.30$, $p = .033$, $d = 1.10$.[15] Interestingly, VVIQ scores also correlated *negatively* with scanning slope on the Dot task for each group ($r = –.52$ for the masters, $r = –.39$ for the novices). Taken together, these findings suggest that vivid visual imagery may help people to scan visual mental images more efficiently, and that chess masters as a group have less vivid imagery than non-masters. Although counter to the stereotype of chess masters, this result recalls the comments made by chess writers such as Fine (1965), who stressed the non-vivid, or more abstract and spatial, aspects of the imagery abilities necessary for chess mastery.

In this experiment we used a direct analogue of a task previously validated to tap image scanning processes to find a substantial and significant difference in the speed with which chess masters and novices are able to visualize the execution of chess moves. This advantage for the masters could not be explained by general superiority in imagery ability[16] or in specific superiority in encoding and responding to chess stimuli, as there were no significant differences on the dot task or in scanning intercept on the chess task. It should be noted that the superiority of the masters was demonstrated in a semantically impoverished context: an incomplete chess position with only five pieces, no kings, and no conceivable matching to long-term memory representations of familiar positions. The results thus support the mental cartoon proposals that visual imagery is an important and

---

[15] A separate experiment (Chabris & McCandless, 1999) obtained VVIQ data from two additional chess masters and 16 additional chess novices who were selected according to the same criteria used for subjects in this experiment. Combining them with the data collected here results in a larger and more reliable difference between groups in imagery vividness: 63/80 for the novices versus 50/80 for the masters, $t(38) = 3.20$, $p = .003$, $d = 1.27$.

[16] This does not foreclose the possibility that experts in other domains posess superior domain-independent imagery abilities; for example, a subgroup of U.S. Air Force fighter pilots may have a flat slope, in milliseconds per degree of rotation, for standard mental rotation ability (Dror, 1992; see also Dror, Kosslyn, & Waag, 1993).

distinct component of expertise in chess and that the visual images processed by experts do facilitate operations specific to their domain of expertise.

The theory does not, however, predict that *any* imagery task involving chess stimuli should be performed better by chess masters than by novices. Only acts of visualization that correspond to specific operations used in the problem-solving process should be enhanced. Future work following up on the results of this experiment should explore the implications of this. For example, Kosslyn's original image scanning demonstration (1973) involved asking participants to focus on one part of a visualized object and then answer questions about a different part, which could be either close by or far away. Time to respond varied with this distance. Since simply shifting attention from one part of the chess board to another is not an operation involved in visualizing the execution of chess moves, masters should scan no faster than novices on a chess analogue of Kosslyn's task that would ask participants to focus on one piece in a chess position and then answer questions about other pieces. (Of course, they may respond more quickly overall, due to greater familiarity with the chess board and pieces.) Such a result would directly support the claim of mental cartoon theory that cartoon representations facilitate not all but *only* important visual operations within the domain.

Another worthwhile followup would be a task similar to the one reported here but with two different conditions: one in which the black-and-white checkered squares are removed from the chess diagram displays, and one in which the checkerboard is retained but the black pieces are replaced by the dots. This would address the question of whether mental cartoons include a base-level structure, without which they do not confer any processing benefits.

Experiment 2: Very Long Term Recognition Memory

The next two experiments turn from the proposed two-dimensional mental cartoon supporting visualization processes in expertise to the proposed distinction between separate stores of pattern and conceptual representations of chess positions. Several recent studies have provided evidence consistent with the idea that chess experts represent knowledge in forms beyond lists or maps of chunks. Freyhof, Gruber, and Ziegler (1992) asked chess players to hierarchically organize the pieces in a chess position and found that masters created groupings with fewer, larger, and more meaningful components than amateurs, and that this difference increased as the position became more typical. Cooke, Atlas, Lane, and Berger (1993), in a variation on the classic study by Bransford and Johnson (1972), compared masters' recall of normal positions with and without prior verbal descriptions of the positions. In the framework proposed here, such a description could be used top-down to activate the appropriate position category, which will in turn facilitate encoding the specific patterns and chunks likely to be contained in the expected type of position. Indeed, the prior verbal description did improve recall performance; unfortunately, no condition with inconsistent verbal descriptions was included for comparison. Kämpf and Strobel (1998) asked chess masters, on each of several trials, to study a solvable tactical position for several seconds and then decide whether a newly presented position was the same or different. Consistent with a separate store of "meaning" information, responses were faster when the change caused the previous solution to the position to no longer work, but only for the strongest group of masters. Malkin (1982, as described by Hartston & Wason, 1985) performed a similar exercise with four players, who after solving a position were asked to recall the locations of pieces that did not participate in the solution. Two grandmasters were accurate; two weaker players were not.

The framework outlined earlier suggests that chess position representations will incorporate both pattern and conceptual information. In a demonstration study in 1998, we asked grandmaster Patrick Wolff to perform de Groot's chess position reconstruction task, with a study time of five seconds per position, and then asked him to explain his approach. He said:

> What I was trying to do was just sort of absorb the position and understand where everything was clustered and try to quickly understand what was "going on," establish some logical connections between things ... the important thing is to try to make everything "make sense."

Wolff's description so far gives roles to both chunk recognition and conceptual analysis of the board configuration. Discussing his performance with one of the particular positions we had asked him to reconstruct, he elaborated:

> I can't remember where this rook is very well because the most important thing is this [bishop] and this [pawn] and this pawn structure, and this [bishop move] is happening, and it's just sort of hard to pay attention to the rest ...

In fact, he recalled the specific pieces he mentioned correctly, but made errors on some of "the rest" that he had difficulty attending to. This explanation suggests an interaction between meaning and patterns, with the meaning not only deriving from the patterns and their interrelationships but also influencing how they are stored and remembered.

The goal of this experiment was therefore to test the proposition that storage of chunk/pattern information and meaning/conceptual information is separated, as diagrammed earlier in Figure 2. The standard short-term recall memory paradigm would not be suitable for this purpose, because only the strongest chess players can extract conceptual information about a position from a brief exposure, and they will perform near ceiling levels on the recall task. To get around these obstacles we decided to ask how well chess masters remember famous chess positions, or specific board configurations that they have been exposed to while learning or studying the game. We reasoned that the long passage of time between study and test, which could be decades, might have different effects on the two types of representation.

We considered the mental cartoon theory, the chunking theory, and a hypothetical "meaning theory," according to which expert memory abstracts away and discards all unimportant information, preserving only the important conceptual aspects of a chess position. We compared these theories by confronting chess masters with famous chess positions that were either correct or altered in one of three ways: by moving one piece to a different location and thereby changing the meaning of the position (a "relevant" change), by moving one piece without changing the meaning, or by moving two pieces without changing the meaning ("irrelevant" changes). The players were to judge simply whether the position was exactly the same as the famous one or whether it had been changed, and to provide this judgment in the form of a confidence level between 0% and 100% that the position was correct. Thus, if they were absolutely certain a position was correct, they might respond 100% yes; if they were absolutely certain a position had been altered, they might respond 0% yes.

Figure 5 illustrates the predictions the three theories would make. The chunking theory, as it represents chess positions with only a context-free map of chunks, would predict that the more pieces were changed, the less confidence subjects would have that the position was correct. The meaning theory would predict equally high confidence for no change and irrelevant change trials, and lower confidence for the relevant changes, because only the latter would cause discrepancies with the conceptual information that has been retained. The mental cartoons hypothesis would combine these two patterns in predicting high confidence in no-change trials, low confidence in relevant-change trials, and intermediate levels in irrelevant-change trials, with confidence in such cases declining with the number of changes until so many changes were made that the meaning was in fact altered.

Method

Participants

24 chess masters volunteered to participate in exchange for either money or chess books. 18 were tested at the 1998 World Open tournament in Philadelphia and were each given $5.00 cash upon completing the experiment; four were tested at the 1998 U.S. Masters tournament in Hawaii and were each mailed two chess books (total retail value $29.90) two weeks later, one was tested at the Au Bon Pain cafe in Harvard Square and was given the same two chess books upon completing the experiment, and one was tested in the laboratory and given three chess books (total retail value $42.85) upon completing this experiment and an unrelated experiment of similar duration. Because the testing was mainly done in public places at times convenient for the participants (e.g., between rounds of a chess tournament), some were allowed to complete the experiment without direct supervision by an experimenter. The July 1998 U.S. Chess Federation (USCF) ratings of the participants ranged from 2200 to 2669 (mean 2361, standard deviation

126); two of the participants were international grandmasters. 23 participants were male and one was female.

Materials and Design

We first selected 30 famous chess positions from famous chess games that appear frequently in chess literature and that we believed would be familiar to our potential participants.[17] We tried to choose the most famous positions that could be altered to create the three variations we required. However, no objective method was used to select the positions, and it is possible that positions equally or even more famous than some of the ones we chose could have been used instead. Still, we asked several participants after the experiment whether they could think of any more famous positions we could have included, and none offered any.

We next created four versions of each position. The *no-change* version was simply the exact configuration of pieces on the board that occurred in the famous game. The *relevant-change* version was the same, except one piece was moved to a different square such that the “meaning” of the position was altered. That is, after the important change, the sequence of moves that was played in the original game would no longer lead to the same outcome, or the strategic theme the position is used to illustrate would no longer apply. The *one-irrelevant-change* version was similar, except that the same piece was displaced to a different square without altering the meaning in this sense. And in the *two-irrelevant-change* version a second piece (of the same color) was displaced in the same way.

---

[17] Sources for these positions are marked with an asterisk (*) in the References section.

Each of the resulting 120 chess positions was diagrammed with the Linares typeface (Alpine Electronics, Powell, WY) and paired with the citation of the game it was taken from or based upon. **Figure 6** shows the caption and four diagrams for one of the test positions.

The 30 famous positions were divided into two sets labeled A and B. Each of these sets were divided into three subsets numbered 1, 2, and 3. Thus, there were six groups of five positions: A1, A2, A3, B1, B2, and B3. Six versions of a printed booklet were created. Each contained all 30 positions, one per page. In the first three booklets, the positions in set A were shown in their no-change versions and the positions in set B were shown in changed versions, and vice-versa for the other three booklets. Within the two groups of three booklets, the three sets of changed positions were rotated through the three change-type conditions in Latin square fashion. Table 1 shows this counterbalancing scheme. With it, six participants would collectively be exposed to all 30 positions in all four versions, while each seeing half of the positions changed and half unchanged.

Booklet	**No Change**	**Relevant**	**1-Irrelevant**	**2-Irrelevant**
1	A	B1	B2	B3
2	A	B2	B3	B1
3	A	B3	B1	B2
4	B	A1	A2	A3
5	B	A2	A3	A1
6	B	A3	A1	A2

Table 1. Assignment of positions to conditions in the design of Experiment 2.

Procedure

Participants were tested individually and were allowed to work at their own pace; however, the instructions encouraged them to spend approximately 30 seconds on each position. They were told that half of the positions would be changed and half would be unchanged from their correct versions, and that all of the changes would be made by moving either one or two pieces from their correct locations to incorrect locations. For each position, they wrote their confidence level on a scale of 0–100% in a space provided next to the diagram and caption. They were also instructed, to write down a correction if they thought the position was changed and they were certain how it had been changed. (Very few corrections were proposed by the participants.) Total time to complete the booklet averaged 25–30 minutes.

Results and Discussion

The results are shown in **Figure 7**. Participants reported highest confidence in the no-change condition (71% ± 5% [95% confidence interval]), lowest confidence in the relevant-change condition (40% ± 9%), and intermediate levels of confidence in the one-irrelevant-change and two-irrelevant-change conditions (59% ± 6% and 59% ± 8% respectively).

We assessed the three theories under consideration by (1) evaluating for each theory a planned contrast whose weights represented the pattern of results predicted by the theory, and (2) evaluating four planned *t*-tests to compare pairs of conditions for which the theories make different predictions. Table 2 shows the numerical predictions ascribed to each theory and the contrast weights computed by subtracting each theory's grand mean from its predictions for the individual conditions. The predicted values were chosen as follows: (1) A prediction of 70% confidence was assigned to the no-change condition as a common baseline, since all three theories predict differences between other conditions and this one; (2) The remaining percentages were chosen to reflect the pattern of these relative differences, with the condition that no contrast weights of 0 could be produced. This condition was imposed because such weights would effectively eliminate conditions with such weights from the comparison; for example, had the chunking theory received a prediction of 70, 50, 50, 30 for the four conditions instead of 70, 50, 50, 40, the resulting contrast weights of 20, 0, 0, –20 would have yielded merely a comparison between the no-change and two-irrelevant-change conditions rather than an evaluation of the complete pattern. Although there are many numerical predictions that satisfy these constraints, the values chosen are reasonable ones, and they capture all of the important distinctions among the three theories.

Theory	**No Change**	**Relevant**	**1-Irrelevant**	**2-Irrelevant**	**Mean**
Chunking	70.0	50.0	50.0	40.0	52.5
(weights)	17.5	–2.5	–2.5	–12.5	0.0
Meaning	70.0	30.0	70.0	70.0	60.0
(weights)	10.0	–30.0	10.0	10.0	0.0
Mental Cartoon	70.0	30.0	50.0	40.0	47.5
(weights)	22.5	–17.5	–2.5	–7.5	0.0

Table 2. Predicted mean confidence levels and associated contrast weights for the three theories being compared in Experiment 2.

The contrast testing the prediction of the mental cartoon theory best accounted for the pattern of results observed and was highly significant, $t(23) = 6.90$, $p = .0000006$, $r =$

.82. The contrasts testing the other two theories were also significant; for meaning, $t(23) = 4.90$, $p = .0002$, $r = .71$; for chunking, $t(23) = 4.26$, $p = .0004$, $r = .66$. This is not surprising, as these three sets of contrast weights are themselves highly intercorrelated. Thus, to directly compare the predictions made for specific conditions by the three theories, we performed two pairs of *t*-tests.

First, the difference between the cartoon and chunking theories is in their prediction of the relationships among the relevant and the irrelevant change conditions. Since relevant changes will cause discrepancies with both meaning and pattern representations of positions, whereas irrelevant changes will mismatch only pattern representations, the cartoon theory predicts that confidence will be lower in relevant-change than in one-irrelevant-change and two-irrelevant-change (30 versus 50 and 40). By contrast, the chunking theory predicts no difference between relevant-change and one-irrelevant change (50 versus 50) and higher confidence in relevant-change than in two-irrelevant-change (50 versus 40). Therefore, we compared the relevant-change condition to each of the irrelevant-change conditions. Confidence was in fact lower in the relevant-change condition than in both the one-irrelevant-change condition (40% versus 59%; $t(23) = 4.24$, $p = .0004$, $r = .66$) and the two-irrelevant-change condition (40% versus 59%; $t(23) = 2.97$, $p = .007$, $r = .53$). In each case, the cartoon theory predicted the results and the chunking theory did not.

Second, the difference between the cartoon and meaning theories is in their prediction of the relationships among the no change and the irrelevant change conditions. The cartoon theory predicts that confidence will be higher in no-change than in one-irrelevant-change and two-irrelevant-change (70 versus 50 and 40), whereas the meaning theory predicts no difference among all these conditions (70 in each case). Therefore, we compared the no-change condition to each of the irrelevant-change conditions. Confidence was in fact higher in the no-change than in both the one-irrelevant-change condition (71% versus 59%; $t(23) = 3.73$, $p = .002$, $r = .61$) and the two-irrelevant-change condition (71% versus 59%; $t(23) = 2.70$, $p = .02$, $r = .49$). In each case the cartoon theory predicted the results and the meaning theory did not. Taken together, these four comparisons show that the cartoon theory predicted better than either of the alternative theories chess masters' behavior when recognizing famous chess positions.

To explore the data further, we computed the correlations between the participants' USCF ratings and their contrast scores for the three theories being compared. The cartoon theory scores correlated more highly with chess skill ($r = .19$) than did the those for the chunking theory ($r = .14$) and those for the meaning theory ($r = .11$), but none of these correlations were significantly different from 0. We also created contrast weights of +75.0, –25.0, –25.0, –25.0 to represent "perfect memory," i.e., confidence of 100% in the no-change condition and 0% in each of the three change conditions. (Not surprisingly, this contrast was significant, $t(23) = 6.39$, $p = .000003$, $r = .80$, since confidence was indeed higher in the no-change than in the change conditions.) The correlation between these contrast scores and USCF ratings was $r = .22$, indicating that better players had superior overall memory for famous chess positions. Finally, we compared the overall mean confidence level of 62% with the 50% level that would have been observed if the participants had successfully used the information that half of the positions were changed and half were not; the difference was significant, $t(23) = 6.60$, $p = .000002$, $r = .81$. This suggests that the input-to-memory comparison is yielding strong enough signals of familiarity to overcome any top-down or base-rate knowledge the participants may have been trying to use when making their judgments.

One aspect of the results that was not consistent with the mental cartoon theory was the apparent equivalence of the one-irrelevant and two-irrelevant change conditions (59% confidence in each case). This flatly contradicts the chunking theory, but it is not inconsistent with the meaning theory. Also supporting the importance of meaning is the finding that the drop in confidence caused by a relevant change (31%, from 71% to 40%) was more than twice as large as the drop caused by irrelevant changes (12%, from 71% to 59%). Collectively, these findings suggest that while the cartoon theory is correct in proposing that both pattern and conceptual information are stored, the latter is stronger, at least for the memories held by chess masters for famous chess positions. It would be of interest to try to adapt this paradigm to see whether the chess memories of weaker players are similarly unbalanced, or whether their access to patterns may be more prominent relative to concepts.

Experiment 3: Very Long Term Recall Memory

The results of Experiment 2 were most consistent with the prediction of the mental cartoon theory that chess masters store at least two distinct types of information about chess positions: piece-location data, probably in the form of chunks, and conceptual data, in an unspecified form. If this is true, it may be possible to measure the different effects of the two types of information in a memory retrieval task.

As discussed previously, chunks as usually understood do not have enough specificity to drive the selection of plans and moves that puts a grandmaster's knowledge into practice, but conceptual information does. What would happen if we asked a grandmaster to recall, with no specific warning or prior study, the famous chess positions used in Experiment 2? If the chunks hold all the meaning information, then pieces that are more important to the position's meaning (i.e., pieces that participate crucially in the continuation the game followed, or pieces that contribute to the strategic principle the position illustrates) should be better represented and recalled from long-term memory than should pieces that are unimportant to the position's meaning. Any errors in recall could be attributed to incorrectly stored or retained chunks, and should result in more or less random perturbations of the correct position after recall. These predictions are consistent with the claim by Ericsson and Delaney (1998) that "selection of relevant information for encoding in long term working memory is not a critical issue for the very best chess players, because the locations of all of the chess pieces are potentially pertinent to the selection of a move for a given chess position." By contrast, if conceptual representations store the meaning of the position alongside context-free chunk representations of specific piece locations, there is no reason to expect a difference in recall of important and unimportant pieces. When a chunk representation is in error, however, we should expect the meaning information to be used to make an "educated guess" about correct piece-location pairs.

We tested these predictions by asking grandmasters Patrick Wolff and Sergey Kudrin to attempt precisely this surprise famous position recall test.

Method

To the 30 positions used in Experiment 2 we added ten more, selected in a similar fashion,[18] for a total of 40 famous chess positions. We presented the ten new ones first followed by the 30 old ones (in the same order as in Experiment 2). Before testing we asked an international master (a holder of the highest chess title below grandmaster) to divide the pieces in each of the 40 positions into approximately equal-size groups of "important" and "unimportant" pieces.

Wolff was tested in a single session of approximately 150 minutes. For the first 30 minutes he completed a short-term chess recall task using ten positions he had never seen before, and discussed his performance on that task (quotations were presented in the introduction to Experiment 2). Then we started the famous positions test. For each position we read him a cue similar to the captions used in Experiment 2, with the addition of either the previous or next move played in the game, so that Wolff could identify which position in the game was being requested. For example, we asked him to set up the position from "Winter–Capablanca, Hastings 1919, White to play (after 15 ... f6)." We

---

[18] Sources for these additional positions are marked with an asterisk (*) in the References section.

also prepared an additional hint to aid memory in cases where Wolff did not recognize the game we were requesting. For example, for the Winter–Capablanca position Wolff needed the hint, which was “White loses because of an entombed bishop,” before beginning to reconstruct the position. The hints did not mention specific piece locations, and were only available before Wolff began to reconstruct the particular position. A standard size chess board and set were used. Wolff indicated when he had finished reconstructing each position (there was no time limit), and the next was begun. A break was taken after every ten positions. The entire session was videotaped, and Wolff’s final reconstruction for each position was recorded from the tape.

Kudrin was tested several weeks later using the same procedure, with two exceptions. First, he was allotted only 90 seconds to reconstruct each position, measured from the time when he indicated he was able to attempt recall. Second, he was not allowed to play out sequences of moves, whether from an earlier position in the same game or from the starting position, to aid his reconstruction of the positions we desired (Wolff had tried this on three occasions). Kudrin’s testing session lasted approximately 120 minutes; quotations from his short-term chess recall task were presented earlier.

Results and Discussion

Wolff attempted to reconstruct 30 of the 40 positions; for the other ten he either did not know of the game at all or knew of it but was unable or unwilling to recall any piece locations. For each position he attempted, we divided his piece placements into correct and incorrect sets using a strict criterion: the correct piece had to be on the correct square for the placement to count as correct, and any other placement was incorrect.

(That is, the only errors scored were errors of commission.) The attempted positions included an average of 22.13 pieces and he placed an average of 18.67 pieces per position. Of these, 15.23 were correct (69% of the total) and 3.43 were incorrect. The important and unimportant piece sets were almost equal in size (11.00 and 11.13 pieces per attempted position respectively), and as shown in **Figure 8**, Wolff placed correctly an average of 7.63 important pieces (69%) and 7.60 unimportant pieces (68%) per position—no difference.

Wolff made a total of 103 individual errors of commission. We scored each error (by consensus of two master-level judges) as either consistent with or contrary to the meaning of the position, with meaning defined in the sense used to create the changes in Experiment 2. By these criteria, as shown in **Figure 9**, Wolff made 68 meaning-consistent errors (66%) and 35 meaning-contrary errors (34%), a significant difference, $\chi^2(1) = 10.57$, $p < .003$. Note that had meaning information been perfectly retrieved, all 103 errors should have been consistent. We suspected that had no meaning information been retrieved (that is, if the chunking theory were governing performance), few of the errors would have been consistent, since filling in gaps in recall by randomly choosing piece-location pairs would usually upset the meaning of a position. To establish a more precise baseline, we randomly selected six of Wolff's errors and for each, we systematically judged whether placing the erroneous piece on every unoccupied square (besides the correct square) would produce a consistent or contrary error. There were 194 piece-square combinations tried, of which 48 (25%) were consistent and 146 (75%) were contrary. Thus, a null hypothesis of an equal number of consistent and contrary errors is conservative with respect to our theoretical predictions.

Kudrin attempted 29 of the 40 positions. The attempted positions included an average of 22.45 pieces and he placed an average of 20.79 pieces per position. Of these, 15.69 were correct (70% of the total) and 5.10 were incorrect. The important and unimportant piece sets were almost equal in size (11.21 and 11.24 pieces per attempted position respectively), and Kudrin placed correctly an average of 8.14 important pieces (73%) and 7.55 unimportant pieces (67%) per position—again, essentially no difference. Kudrin made a total of 148 individual errors of commission, of which 92 (62%) were meaning-consistent and 56 (38%) were meaning-contrary, a significant difference, $\chi^2(1) = 8.76$, $p < .007$. Thus the performance of Kudrin, under somewhat stricter conditions, was almost identical to that of Wolff.

The consistent findings of no difference in recall between important an unimportant pieces, combined with significantly more meaning-consistent than meaning-contrary errors, are as predicted by the theory that grandmaster representations of chess positions incorporate separate components of chunk and conceptual information. Furthermore, the verbal reports of the participants during the experiment also support this idea. For example, Kudrin only attempted to place two pieces out of sixteen when recalling the position from a 1925 game between Richard Reti and Alexander Alekhine, and in explaining his poor performance he said: “I can’t do much. I will put a couple of pieces, but that’s about it. I don’t remember any of the mechanism of how everything was connected.” He was referring to the fact, which he knew, that most of the pieces were involved in a long and complicated combination, but he could not recall *how* they were involved, conceptual information which would have helped him to determine their precise identities and locations.

Taken together, the results of Experiment 2 and 3 clearly support the distinction, incorporated in the mental cartoon theory, between pattern and conceptual knowledge in the domain-specific memory representations of chess masters. The chunking theory was not able to fully account for the results: it did not predict the difference between relevant and irrelevant changes in Experiment 2, and it did not predict the predominance of meaning-consistent errors in Experiment 3. The meaning theory was also not fully supported: it did not predict the difference between no-change and irrelevant-change conditions in Experiment 2, and it did not predict the equal recall of important and unimportant pieces in Experiment 3.

Experiment 4: Benefits of Thinking Ahead

Advocates of the chunking theory, according to Gobet & Simon (1996b, p. 53), "have proposed that recognition, by allowing knowledge to be accessed rapidly, allows the slower look-ahead search to be greatly abridged or even dispensed with entirely without much loss in quality of play." The implication is that beyond a certain minimum time to make a decision, additional thinking time does not improve decision quality. This follows naturally from the lack of a role for longer-term conceptual knowledge development in the chunking theory, under which only small, readily-identified patterns are the locus of knowledge that supports skilled performance. Calderwood, Klein, and Crandall (1988) found evidence for the primacy of recognition as a component of skill by asking grandmasters to compare the quality of moves played in tournament games (average decision time, 135 seconds per move) with moves played in blitz games (average decision time, 5 seconds per move). The moves played by master subjects did not differ as much in these quality ratings between the two conditions as did the moves played by weaker subjects.

In an attempt to improve on the subjective aspect of this method, Gobet and Simon (1996b) carried out a clever study to test whether pattern recognition is more important than thinking ahead in determining chess playing strength. They examined a sample of game results of the current world chess champion, Garry Kasparov, playing under conditions where he had less time to think than his opponents. This happened during exhibitions in which he played games against several strong players simultaneously, under timed conditions. For example, in January 1992 he played four members of the German national team at once, and thus had about one-fourth as much time to think about each game as his opponent did. From the 56 game outcomes they collected, Gobet and Simon estimated that Kasparov's strength in such conditions was not notably lower than his playing strength under normal tournament time limits. They therefore concluded that recognition processes, which occur nearly instantaneously, probably dominate planning and look-ahead processes, because otherwise Kasparov would have performed much worse than he did.

This would seem to be persuasive evidence—if the world champion's play does not suffer from having only one-fourth as much time for thinking as he usually does during a game, how can the thinking be very important? However, Gobet and Simon failed to recognize the variability in their estimate of Kasparov's strength under these time limits, and therefore could not describe the uncertainty associated with their conclusion. Furthermore, the Elo linear performance rating formula they used to compute an estimate of Kasparov's strength has poor statistical properties (e.g., it is not a consistent estimator, and is only intended as a quick approximation to more precise methods), so it should not be relied upon. We reanalyzed their data using maximum likelihood estimation to obtain more principled inferences about Kasparov's playing strength, and to calculate its standard error.

After correcting a minor error in Gobet and Simon's data set,[19] we carried out this analysis on the 56 game outcomes. The maximum likelihood estimate is 2594 with a standard error of 53.6. Thus, a 95% confidence interval for Kasparov's true rating over

---

[19] Kasparov's score in his match against the United States junior team was actually 0.5 points worse than that reported by Gobet and Simon.

the 56 games is 2489–2699. This interval estimate suggests a different conclusion than the one reached by Gobet and Simon, who obtained a point estimate of 2646. Although it is clear that Kasparov played worse than his rating range of 2700–2790 during the same time period (July 1985 – July 1992) would indicate, there is not enough information to determine how much worse. For example, the confidence interval suggests that under simultaneous conditions, Kasparov may be playing at a rating level 200 or more points below his normal tournament strength.

A rating difference of 200 points is significant because it predicts a 3–1 victory margin for the superior player in a match, no matter where along the absolute rating scale the two players fall (because the rating scale is logarithmic; Elo, 1986; cf. Glickman, 1995; for caveats see Glickman & Jones, in press). This would be considered a decisive result; for comparison, all recent world championship matches have been decided by much smaller margins. Thus, if Kasparov lost 200 points of strength under clock-simultaneous conditions, it would be fair to conclude that the lost thinking time affected his play significantly. It is interesting to note that research in computer chess (e.g., Thompson, 1982; Condon & Thompson, 1983; Newborn, 1985; Hsu, Anantharaman, Campbell, and Nowatzyk, 1990; Hyatt & Newborn, 1997; for a review, see Heinz, 1998) has equated a 200-point rating advantage to the approximate benefit derived from searching one ply (one move for one side) deeper in the game tree, and that this additional search typically increases the time spent by a factor of 4–6. Thus, under the same time constraints, Kasparov suffers the same performance decrement as a typical chess-playing computer, which of course has much less knowledge and poorer pattern-recognition ability than he does, and derives most of its skill from efficient tree-searching. One can at least conclude from this similarity that Kasparov and other grandmasters derive a significant part of their skill from slow thinking processes as opposed to rapid perceptual processes.

To further explore the value of thinking ahead, we decided to study not simply the game outcomes and mean performance level of grandmasters under different time conditions, but the actual quality of the individual decisions of which an entire chess game is composed. We analyzed “blunders,” or gross errors, made by many of the world’s top players in a specially created database of games. From 1993 to 1998 an annual tournament has been held in Monaco in which 12 grandmasters play one another once each under “blindfold” and “rapid” (sighted) conditions, in each case at a fast time limit of 25 minutes per player for the entire game. This controls for differences in players, dates, time limits, and other factors that would be present if we simply sampled blindfold and sighted games from disparate events. These six tournaments comprise 396 pairs of games. Following a case-control design, we matched to each pair the most recent game played in another event by the same two players under sighted, slow time conditions (known as “classical” chess: the traditional three minutes per move instead of rapid chess at 25 minutes per game). We excluded games ending in less than 20 moves; in 40 cases there was no matching game in our database (probably because the two players had not played each other outside of the Monaco tournament), so we used instead a game played between two other players who had competed at Monaco. Thus, the classical game selection included 396 games played among the players who had competed in the Monaco series. All game scores were obtained from the *ChessBase 7* software package (ChessBase, Hamburg, Germany).

To identify all the significant errors made in the resulting set of 1188 games we processed them through the game analysis facility of the *Fritz 5* chess playing program (ChessBase, Hamburg, Germany) running on a Dell Dimension XPS 233 MHz Pentium II desktop computer system.[20] The program analyzed every move played with a nominal 10-ply exhaustive search and reported as "candidate blunders" all cases in which the move played was evaluated as 1.5 pawns or more worse than the program's choice for the best move.[21] That is, a move was considered a possible blunder if *Fritz 5* determined that the best move in the given position had a minimax value at least 1.5 pawns better than the move actually played in the grandmaster game; whether the blunder was exploited during the game was irrelevant to this analysis. The 1.5-pawn cutoff was not chosen arbitrarily; an advantage of this size is believed by computer chess researchers to be sufficient to win the game, and according to Hartmann (1989), who computer-analyzed 832 games (comprising 62,965 moves) by human masters, it "could very well be the threshold beyond which a game is theoretically won." To further refine our set of blunders, we used a conservative criterion to exclude all errors that did not alter the probable outcome of the

---

[20] We plan to replicate this analysis in the future using the *Deep Blue* computer system of IBM Research, with whom we are collaborating on this ongoing project. *Deep Blue* (Hsu, Anantharaman, Campbell, & Nowatzyk, 1990) is noteworthy for having defeated Garry Kasparov in a match (Goodman & Keene, 1997) and is generally considered the strongest chess computer ever developed. The *Fritz 5* program used here is one of the strongest commercially-available microcomputer programs according to the Swedish Computer Chess Federation rating list, though its decisions will naturally be less reliable than those of *Deep Blue.* As chess computers are at their worst in situations involving subtle distinctions and long-range judgments, we have used a high threshold for considering a suboptimal move a "blunder" in this study in order to minimize the rate of false positives.

[21] Chess computers judge chess positions using many criteria, usually several hundred or thousand, but they express them all on a common scale in terms of pawn units. Having one more pawn than one's opponent in an otherwise perfectly balanced position would correspond to an evaluation of +1.0 pawns. A nominal 10-ply search refers to the depth, in moves, to which the program exhaustively analyzed every position in every game in the database. The search is often extended deeper when it reaches an unclear or unstable position.

game. That is, if one side made an error of 1.5 pawns or more, but even after that error the same side still had an advantage of *3.0 pawns or more*, the error was excluded. The resulting set was considered to include only “true blunders.”

	Games	**Moves**	**CB**	**TB**	**TB_{EXP}**	**TB/KM**	**B_{MEAN}**	**B_{SE}**
Classical	396	17518	252	176	229	10.0	2.66	0.11
Rapid	396	19408	447	266	253	13.7	3.15	0.13
Blindfold	396	18156	424	277	237	15.3	3.08	0.12
	1188	*55082*	*1123*	*719*	*719*	*13.0*	*2.96*	

Table 3. Results of Experiment 4, comparing blunders by grandmasters in classical, rapid, and blindfold playing conditions. *Games:* total games in database played under each condition. *Moves:* total moves played under each condition. *CB:* candidate blunders (see text). *TB:* true blunders (see text). *TB_{EXP}:* true blunders expected in each condition assuming that true blunders were distributed among conditions according to the total moves in games played in each condition. *TB/KM:* true blunders per thousand moves. *B_{MEAN}:* average magnitude, in pawn units, of the blunders in each condition (after removing blunders that allowed checkmate, which have infinite magnitude). *B_{SE}:* standard error of B_{MEAN}.

Table 3 shows the results. In the 1188 games analyzed we found 1123 candidate blunders, of which 719 were true blunders. Contrary to the stereotype of chess grandmasters as people who may make subtle errors but never serious mistakes, we found one true blunder per 1.65 games. Moreover, 31 of the true blunders (4.3%) involved permitting checkmate, immediately ending the game (though often after a forced sequence several moves long). Broken down by game type, there were 176 true blunders in the classical games, 266 in the rapid games, and 277 in the blindfold games. To properly test our hypothesis that blunders should be more frequent in rapid games than classical games, we divided the number of true blunders in each condition by the total number of moves played in the games in that condition. Grandmasters made 10.0 blunders per 1000 moves in classical games, 13.7 in rapid games (an increase of 37%), and 15.3 in blindfold games. It is interesting to note that the advantage of additional time (classical versus rapid, 3.7 blunders per 1000 moves), is over twice as large as the advantage of being able to see the chess board and pieces (rapid versus blindfold, 1.6 blunders per 1000 moves). This difference cannot be explained by a ceiling effect, since play can be *much* worse than grandmaster blindfold play: Theoretically, *every* move could be a blunder, and it is not hard to envision 10–15% blunders, or about ten times the proportion found here.

To assess the statistical significance of these differences, we calculated the expected number of true blunders in each condition under the null hypothesis that blunders would be distributed among conditions in proportion to the total number of moves played in each condition (see Table 3). The frequency of blunders differed significantly among conditions, $\chi^2(2) = 19.52$, $p = .00006$. Furthermore, the magnitudes of the blunders made in each condition differed, as shown in Table 3; crucially, blunders were less severe under classical conditions than under rapid conditions, $t(419) = 2.71$, $p = .008$, $r = .13$. Thus, grandmasters apparently make significantly fewer and smaller mistakes when they have additional time to think ahead during a game, contrary to the conclusions of Gobet and Simon (1996b) and to the prediction extrapolated from the chunking theory's emphasis on fast pattern recognition over slow thinking processes as the source of expertise.

We cannot yet compare directly the 37% increase in blunders we found here to the rating penalties calculated earlier from the Kasparov game sample. One way to do so in the future would be to derive a function to map frequency of blunders onto Elo-scale chess ratings; we plan to do this via computer analysis of a very large database of games by players of all skill levels. At this point, however, it should be noted that our conclusion of a substantial benefit from thinking ahead is based on analyzing all 55082 moves played in 1188 games among grandmasters, whereas Gobet and Simon used only the final outcomes of 56 games played by a single player, many of them against non-grandmaster opponents.[22]

The results of Experiment 4 do not support the claim of Gobet and Simon that pattern recognition processes play a dominant role in top-level chess expertise. Instead, they point to a significant contribution from thinking ahead, consistent with the model proposed here. It could still be the case that experts differ from novices *more* in pattern recognition than in other processes, but it cannot be the case that experts do not suffer from a reduction in decision time.[23]

---

[22] It is important to note that the method used by Gobet and Simon (1996b) suffers further imprecision from the fact that all of Kasparov's opponents in the exhibitions were weaker, often substantially, than he was. This is natural, as he was the world champion during the period under consideration. Nevertheless, and without addressing the accuracy of the underlying Elo rating system model (see Glickman, 1995; Glickman & Jones, in press), a more precise rating estimate could be obtained by considering a range of opposition on both sides of, or at least closer to, the target player's own strength. The method we used here, since it compares the behavior of players who compete against one another rather than a separate, weaker group, does not suffer from this limitation.

[23] The discrepancy of our results with those of Calderwood, Klein, and Crandall (1988) may lie in the form of measurement—their subjective ratings were in fact unable to detect a significant difference in move quality between masters and amateurs under normal tournament conditions (mean move quality ratings, 2.98 versus 2.97 respectively on a 1–5 scale), despite the very large difference in externally-rated (Elo system) playing strength between these two groups.

Hemispheric Specialization

The foregoing discussion has characterized mental cartoons in terms of traditional cognitive theory, leaving open the questions of how and where they are implemented in the brain. The former is premature at this point, but the latter can be attacked indirectly by research on the neuropsychological basis of chess skill. If the development of a mental cartoon is the critical information-processing factor that turns novices into experts in chess, then asking what brain areas are involved when skilled players perform chess-related tasks may suggest where we might find mental cartoons in the brain.

Cranberg and Albert (1986) gathered three main sources of evidence to address the hemispheric basis of chess skill. Reviewing an EEG study finding increased right-hemisphere activity during blindfold play by a master, conducting a survey to show that left-handedness is more frequent among chess players than among non-players, and discovering in the literature several cases in which left-hemisphere brain damage (with aphasia) did not impair chess playing ability, Cranberg and Albert concluded that the right hemisphere is critically involved.[24] This finding is supported by studies by Frydman and Lynn (1992), who report that young tournament chess players have above-average performance IQs but normal verbal IQs, and Horgan and Morgan (1990), whose skilled young players scored above age norms on Raven's Progressive Matrices test (which

---

[24] One of the most popular neurological cases, the "man who mistook his wife for a hat" described by Sacks (1985), appeared to suffer some form of visual agnosia and/or prosopagnosia, possibly from right-hemisphere damage, but was still able to play blindfold chess (and defeat his doctor). However, Sacks did not see whether the patient could play chess *with* sight of the board, nor did he determine the precise locus of the brain lesion involved. It is possible that the sort of spatially-demanding visualization involved in blindfold chess does not depend critically on the occipital-temporal areas that may have been damaged in this case.

measures abstract and spatial reasoning abilities rather than verbal ones). Experiments 5 and 6 were designed to investigate hemispheric specialization for the components of chess skill common to all theories of expertise: chunking and pattern recognition.

Computational analyses of the problems that must be solved by any artificial or biological perceptual machine (Marr, 1982) suggest that the human visual system requires both a rule-following “default” and a rule-violating “override” mode of parsing to efficiently process the varied stimuli we encounter in everyday life. The visual system will employ as its default the parsing strategy most likely to yield meaningful groupings of stimulus elements, but it will override this strategy (presumably with some sort of attentional control) when necessary to solve a problem at hand (Van Kleeck & Kosslyn, 1989). For example, boundaries between object parts usually occur along edges, or sets of collinear points, so the visual system should usually group such points together when organizing an image. However, when two distinct object parts of similar color and texture are juxtaposed in the image, the visual system must overcome its tendency to combine the parts into a single whole.

Divided visual field studies have shown that the two cerebral hemispheres differ in their abilities for perceptual organization (Van Kleeck, 1989; Van Kleeck & Kosslyn, 1989). Specifically, the right hemisphere has been shown to perform better than the left at parsing according to Gestalt principles (Wertheimer, 1938), such as the laws of collinearity and similarity involved in the above example. The left hemisphere outperforms the right, however, when a parse is demanded which violates one or more of these principles. The contrast between the abilities of the two hemispheres to perform the two types of parsing tasks is further evidence that the brain uses separate processes for parsing in the two different modes. However, the studies which have shown these lateral differences have all used semantically neutral stimuli such as capital letters (Van Kleeck, 1989) or meaningless combinations of line segments (Van Kleeck & Kosslyn, 1989). In a domain like chess, specific interrelationships among elements are essential for solving problems. For optimal performance, the visual system must encode stimuli in ways that make explicit these useful relationships rather than the simple perceptual regularities.

The default/override framework of Van Kleeck and Kosslyn (1989) can be extended in two ways to predict hemispheric differences in chunking ability. At first glance, it might seem that parsing according to the chess chunking rules described above should usually require override of the visual system's default parsing rules, since the chunking rules may generate groupings that would violate Gestalt principles. The chess position diagrammed in **Figure 10** is an extreme example of this conflict. The highlighted pieces in set A would be chunked together by most chess masters despite the fact that such a grouping violates the Gestalt principles of proximity, similarity, and collinearity; by contrast, the pieces in set B would probably not be chunked together even though they are proximal, similar in shape, and collinear. If override is therefore necessary to impose the domain-specific organization on the stimulus—the "pure override" hypothesis—and the left hemisphere is superior at overriding default schemes of perceptual organization, then chess positions should be chunked and remembered better by the left hemisphere.

Alternatively, the rules of perceptual organization for a particular domain could supplant the preexisting domain-independent rules and "become" the visual system's default rules for use within the context of that domain. By this view, one component of skill acquisition would be the learning of new parsing rules that enable more efficient and relevant encoding of stimuli peculiar to the new domain. In the case of chess, as a player's skill increases with practice, his visual system will adapt to "see" a chess position as a collection of meaningful groupings of pieces, using as its defaults the chunking rules. Accordingly, by this view—the "acquired defaults" hypothesis—chess positions should be chunked and remembered better by the right hemisphere.

The limited neuropsychological evidence reviewed earlier suggests that the right hemisphere is somehow critical for the development and manifestation of chess skill. This is clearly consistent with the acquired defaults hypothesis, that the right hemisphere's superiority at acquiring and applying sets of default parsing rules enables it to better organize chess positions for memory and other types of processing necessary to play the game. On the other hand, in cases when chunking is not effective, the left hemisphere may become involved in assigning "meaning" to stimuli (cf. Gazzaniga, 1985), though this is not likely to be as successful.

To test the alternative hypotheses, we performed two divided visual field experiments on groups of 16 chess masters. The use of this paradigm was possible because reliable skill differences in chess memory have been found with presentations as brief as 150 ms (e.g., Ellis, 1973), a duration well below the upper limit of 200 ms recommended for divided visual field studies (Beaumont, 1982). Experiment 5 tested recognition memory for chess positions, comparing performance on realistic and randomly-generated positions. Experiment 6 focused on the question of parsing by testing the ability to determine whether the piece configurations in briefly exposed 4 x 4 fragments of chess positions were contained within previously studied complete 8 x 8 realistic positions.

Experiment 5: Asymmetries in Recognition Memory

Chunking was originally measured with a free-recall test of memory and a task of copying chess positions in free view onto an empty board (Chase & Simon, 1973a, 1973b). Since, like recall memory, recognition memory for chess positions correlates well with chess playing ability (Goldin, 1979), we can assume that testing it will engage the same domain-specific memory skills.

If the pure override hypothesis is correct and chunking is the imposition of a nondefault perceptual organization on the stimulus, then the default/override framework predicts that subjects should perform better when chess positions containing familiar patterns are presented briefly in the right visual field and are transmitted initially to the left hemisphere. This is because the left hemisphere seems superior at overriding default parsing systems such as the Gestalt laws (Van Kleeck & Kosslyn, 1989), and subjects should respond more quickly or accurately when stimuli are presented initially to the hemisphere that is better at using the information (Beaumont, 1982). Randomly generated positions, however, will contain few patterns that obey either the chess chunking rules or the Gestalt laws. In this case, the left and right hemispheres will better encode different sections of the position depending on how their component pieces can best be grouped or whether meaning can somehow be assigned. Therefore, with random positions there should be little or no asymmetry in performance between left- and right-hemisphere presentations.

Method

Participants

16 male chess masters volunteered to participate and were paid at an hourly rate for their time. Their January 1989 U.S. Chess Federation (USCF) ratings ranged from 2149 to 2467 (mean 2305). All were right-handed as measured by the Edinburgh Handedness Inventory (Oldfield, 1971) and had normal or corrected-to-normal vision. Although some of the participants were familiar with the general idea of chunking from reports in the chess literature, none were aware of the specific purposes or predictions of this experiment.

Stimuli and Apparatus

Pictorial diagrams of the sort that appear in chess books and magazines, and similar to those used in previous studies of chess memory (e.g., Charness, 1976), were drawn with a typeface similar to those used in Experiments 1–3, for 24 chess positions. 16 were classified as “normal” positions, and were selected from games played in the four world championship matches of 1984–85, 1985, 1986, and 1987. Half of these normal positions were endgames (defined as having 16 or fewer pieces on the board) and half were middlegames (17 or more pieces). The remaining eight were classified as “random” positions, and were created by choosing 16 pieces at random from the complete set of 32 and distributing them randomly on the board. This procedure virtually guaranteed that the random positions contained configurations of pieces that could never occur in an actual game.

The positions were divided into two sets of eight (two sets of four for the random task), A and B. Half of the subjects memorized set A as the targets and saw set B as the distractors, and half memorized set B as the targets and saw set A as the distractors. To ensure that the pieces in each position would appear equally often in the more and less acute (central and eccentric) portions of each visual field, each position was diagrammed from both sides of the board. (For a normal position, this represents viewing the board separately from the point of view of the White and Black players.) Examples of normal and random positions, drawn in a font similar but not identical to that actually used for the stimuli, are shown in **Figure 11**.

The diagrams were square, with each side measuring 7.7 cm, and subtended 8.75 degrees of visual angle when viewed at a distance of 50 cm (enforced with a chin rest). Stimuli were presented with the MacLab program (Costin, 1988), which recorded responses and latencies, on an Apple Macintosh Plus computer with a Polaroid CP-50 polarizing screen filter to reduce glare.

Procedure

Subjects were tested individually in single sessions which varied in duration from two to nearly four hours. A session consisted of two tasks—tests of recognition memory for the normal and random positions—with a short break in between. Half of the subjects were tested first on the normal positions, and half were tested first on the random positions. The two tasks were identical except for the different stimuli and as otherwise noted below. Each task consisted of a memorization phase followed by a testing phase.

In the memorization phase, subjects memorized the target set of eight positions (four for the random task) by studying their diagrams and reconstructing them with a standard chess set and board. Going through the set several times in the same pseudorandom order, alternating between the two views of each position, they continued until they had correctly recalled each position three times.

The testing phase consisted of 128 experimental trials (64 for the random task), preceded by 16 practice trials (8 for the random task) selected from the experimental trials and including each target and distractor position once. At the beginning of each trial, the subject focused on a fixation point, which remained in the center of the screen for 1000 ms; this was followed by one of the diagrams, which appeared for 180 ms with its inner edge 1 cm (1.15 degrees of visual angle) to the left or right of the fixation point. Subjects were instructed to decide as quickly and accurately as possible whether the diagrammed position was one they had memorized earlier. The fixation point remained on the screen while they made their response, and the next trial began immediately thereafter. Half of the subjects indicated “yes” responses by pressing the “.” (period) key with their right hand and “no” responses by pressing the “z” key with their left hand on the computer keyboard; the other subjects responded “yes” with their right hand and “no” with their left hand.

The experimental trials were divided into two blocks of 64 (32 for the random task), which were presented sequentially. In each block, the 16 (8 for the random task) positions were presented once viewed from each side of the board (from the white and black players’ viewpoints) in each visual field. Trials were arranged in different pseudorandom orders in each block, with the constraints that no position appeared twice before all appeared once, three times before all appeared twice, and so on, and no correct response (yes or no) or visual field of presentation (left or right) appeared more than three times in a row. Half of the subjects received the trials as described above; the others received them with left- and right-visual-field presentations reversed.

Results and Discussion

We considered only "yes" trials, those in which the probe chess position was one of those the subject had memorized, because we were interested in the processes of encoding and accessing chess-specific long term memories. Since "no" trials will not result in such memories being matched, we cannot be certain that they engage the processes we are studying.

Error rates and mean response times were subjected to separate analyses of variance with identical designs. The factors were hemisphere (left or right), task (normal or random), order of tasks (normal first or random first), and target/distractor combination (A/B or B/A). In the error rate analysis, all responses were included; in the response time analysis, trials were eliminated from consideration if either the incorrect answer was given or the time exceeded twice the mean of the other trials in that cell of the design for the particular subject. All effects and interactions not noted were nonsignificant, $p > .10$ in all cases.

The most interesting results are illustrated in **Figure 12**. As expected, there was an interaction between task and hemisphere, though only marginally significant, $F(1,12) = 3.73$, $p < .08$, $r = 49$. Contrasts were used to examine more closely the response time differences between hemispheres for each position type. Responses to random positions were faster for left hemisphere trials (1372 ms) than for right hemisphere trials (1489 ms), $F(1,12) = 6.37$, $p < .05$, $r = .59$. Responses to normal positions appeared to be faster for right hemisphere trials (1228 ms) than for left hemisphere trials (1290 ms), but the effect was not significant, $F(1,12) = 1.79$, $p > .20$, $r = .36$. Therefore, the trends for both tasks were opposite those predicted by the pure override hypothesis.

The presence of a main effect of task, with subjects responding faster to normal than to random positions, $F(1,12) = 7.43$, $p < .02$, $r = .62$, indicates that the recognition test did engage processing and/or representations specific to the chess domain. Additionally, there was an interaction between task and order, $F(1,12) = 6.61$, $p < .03$, $r = .60$, because subjects tested on random positions first performed better on normal positions than did subjects tested on normal positions first.

For error rates, the interaction between task and hemisphere was nonsignificant, $F < 1$. Contrasts revealed that there were no significant differences in accuracy between hemispheres for each position type, so there were no speed-accuracy tradeoffs. Error rates on random positions tended to be lower for right hemisphere trials (26.6%) than for left hemisphere trials (30.5%), $F(1,12) = 2.04$, $p > .10$, $r = .38$. Error rates on normal positions tended to be lower for right hemisphere trials (18.6%) than for left hemisphere trials (22.7%), $F(1,12) = 2.39$, $p > .10$, $r = .41$.

The main effect of task was observed again, with subjects performing more accurately with normal than random positions, $F(1,12) = 6.10$, $p < .03$, $r = .58$. There was a marginally significant effect of order, $F(1,12) = 4.58$, $p < .06$, $r = .53$, with subjects tested on random positions first making fewer errors overall. Finally, there was an interaction between order and hemisphere, $F(1,12) = 5.70$, $p < .04$, $r = .57$, with the right hemisphere more accurate overall than the left when the normal task was first (26.6% versus 36.1%), but little difference when the random task was first (18.6% versus 17.0%).

The consistent main effect of position type suggests that chess-specific representations are involved, since normal positions were recognized faster and more accurately than random positions (replicating previous findings; Goldin, 1979). However, a left-hemisphere advantage for normal positions was *not* found, so the results provide evidence against the pure override hypothesis. Moreover, the fact that the task by hemisphere interaction in response times approached significance *with trends opposite to those predicted* indirectly supports the acquired defaults hypothesis.

However, another account is possible, resting on the evidence for a general right-hemisphere superiority in memory of complex visual stimuli like faces (e.g., Milner, 1968), which could extend to chess positions as well. If the right hemisphere remembered both normal and random positions better than the left, there might still be no difference in performance on the random task if it were too difficult (indeed, the mean error rate on the random task was 24.6%). Another potential complication is that some subjects reported in debriefing that they were not consciously aware of “seeing” the entire probe position when it flashed for 180 ms, raising the possibility that they made their decisions on the basis of a few features of the stimuli rather than a complete encoding.

Experiment 6: Asymmetries in Visual Parsing

Experiment 6 directly tested the acquired defaults hypothesis by focusing on the issue of parsing, and addressed the problems with Experiment 5 by using simpler probe stimuli and seeking a left-hemisphere advantage for one condition (rather than the absence of a difference, which could still be indicative of a ceiling effect). We adapted the design used by Van Kleeck and Kosslyn (1989) with an embedded figures task, which required subjects to determine whether a small line drawing was embedded within a larger figure. They reasoned that if the right hemisphere is superior at applying default parsing rules, then it should perform better than the left hemisphere when the constituent was a "good part" parsed from the whole in accordance with the Gestalt laws of perceptual organization, but that for "bad parts" that violate those rules, the left hemisphere should perform better than the right. We adapted this idea to the chess domain, in which the default rules are those that group sets of pieces into chunks, good parts correspond to fragments of chess positions that contain only pieces from a single chunk, and bad parts correspond to fragments containing pieces drawn from two separate chunks. Our acquired defaults theory thus predicts a right-hemisphere advantage for single-chunk fragments and a left-hemisphere advantage for multiple-chunk fragments.

Method

Participants

16 male chess masters volunteered to participate and were paid at an hourly rate for their time. Their January 1989 U.S. Chess Federation (USCF) ratings ranged from

2149 to 2467 (mean 2302). 14 of the masters had participated in Experiment 5 (which took place four months before Experiment 6); the two new participants were also right-handed with normal or corrected-to-normal vision. None of the participants were aware of the specific purposes or predictions of this experiment.

Stimuli and Apparatus

Four chess players, three masters and one player judged by an experimenter familiar with his skill to be of equivalent strength, participated in a pilot study to select the test stimuli. (These players did not participate later in the main experiment.) This task was similar to the memorization phase of Experiment 5, and consisted of 48 trials presented in a single pseudorandom order. Stimuli were photocopies of 48 pictorial diagrams, published in *New In Chess* magazine, of chess positions from international tournament games played in 1987 and 1988. 24 of the positions were endgames (16 or fewer pieces) and 24 were middlegames (17 or more pieces). In each trial, subjects studied a diagram for 30 seconds and then had 15 seconds to reconstruct the position from memory with a standard chess set and board. During the reconstruction, the experimenters recorded the order in which pieces were placed on the board and the groups which were placed together (suggesting chunk boundaries). Videotapes of each session were reviewed to verify these data.

The 16 endgames and 16 middlegames on which the pilot subjects best performed this recall task were used in the main experiment. Study diagrams (8 x 8) were created for each of the 32 positions with the typeface used in Experiment 5. Each position was depicted only from the white player's point of view, as is customary in chess publications.

From each position two sets of pieces were chosen such that one set included pieces drawn from a single chunk and the other set included pieces drawn from two separate chunks. Pieces were assigned to the same chunk if a majority (or in a few difficult cases, a plurality) of the pilot subjects grouped them together when reconstructing the position. The single-chunk and multiple-chunk sets chosen for a position had the same number of pieces, either two or three. 64 "fragments" were created by arranging the pieces from each set on 4 x 4 portions of chessboards. The pieces were placed on squares of the same colors and in the same relative locations as they appeared in the 8 x 8 position from which they were taken. Note that a fragment was *not* a complete 4 x 4 excerpt of the position from which its pieces were taken. **Figure 13** shows an example of a study position and the fragments associated with it.

The study diagrams were identical in appearance to the stimuli used in Experiment 5. The fragments, rendered in the same font as the diagrams, were square, with each side measuring 3.8 cm, and subtended 4.35 degrees of visual angle when viewed at a distance of 50 cm (enforced with a chin rest). Stimuli were presented and response data were collected with the apparatus used in Experiment 5.

Procedure

Each participant was tested individually in a single session of approximately one hour, including eight practice trials followed by 256 experimental trials. At the beginning of each trial, a large exclamation point appeared in the center of the computer screen for

400 ms to inform the participant that the next chess position was about to appear. The screen was then blank for 250 ms. The diagram then appeared in the center of the screen for 3000 ms, during which time the participant was to memorize the position. The diagram was replaced by a mask for 500 ms, followed by a centered fixation point for 250 ms, and then a fragment with its inner edge 1 cm (1.15 degrees of visual angle) to the left or right of the fixation point for 150 ms. The participant was to decide as quickly and accurately as possible whether the pieces in the fragment appeared on the same squares relative to one another in the full position studied. The fixation point remained on the screen while he made his response, after which the mask reappeared for 500 ms before the next trial began. Participants responded with the computer keyboard in the same way as in Experiment 5—half responded "yes" with their right hand, the other half with their left hand.

The experimental trials were divided into four blocks of 64, presented sequentially, with subjects allowed a brief rest after completing each block. Within each block, the 32 study positions appeared twice, with no position appearing a second time before each had appeared once. Furthermore, each block contained equal numbers of single- and multiple-chunk fragments, left- and right-visual-field presentations, and yes and no responses. Within the 256 total trials, each of the 128 position/fragment pairs appeared once in each visual field. Trials were arranged in a single overall pseudorandom order, with the constraints that no position type, fragment type, correct response, or visual field of presentation appeared more than three times in a row. Half of the subjects received the trials as described above; the others received them with left- and right-visual-field presentations reversed.

To ensure that the subjects understood the task, especially the possibly counterintuitive criteria for responding “yes” and “no,” they were shown examples of the stimuli on paper and tested on a set of eight practice trials, none of which appeared later in the experiment.

Results and Discussion

As in Experiment 5, we considered only “yes” trials, those in which the fragment displayed included a part of the position studied. Error rates and mean response times were subjected to separate analyses of variance with identical designs. The factors were hemisphere (left or right), fragment type (single- or multiple-chunk), and position type (endgame or middlegame). In the error rate analysis, all responses were included; in the response time analysis, trials were eliminated from consideration if either the incorrect answer was given or the time exceeded twice the mean of the other trials in that cell of the design for the particular subject. All effects and interactions not noted were nonsignificant, $p > .10$ in all cases.

The most interesting results are illustrated in **Figure 14**. As predicted, there was an interaction between fragment type and hemisphere, with single-chunk fragments processed faster when presented initially to the right hemisphere, and multiple-chunk fragments processed faster when presented initially to the left hemisphere, $F(1,15) = 8.47$, $p < .02$, $r = .60$. Contrasts revealed that subjects responded faster to multiple-chunk fragments presented initially to the left hemisphere (1915 ms) than to the right hemisphere (2074 ms), $F(1,15) = 24.73$, $p < .001$, $r = .79$. For single-chunk fragments, there was a nonsignificant trend in the predicted direction, towards faster responses by the right hemisphere (1643 ms) than by the left hemisphere (1670 ms), $F < 1$. There was also a main effect of fragment type, with single-chunk fragments processed faster than multiple-chunk fragments, $F(1,15) = 34.32$, $p < .0001$, $r = .83$. As in Experiment 5, this validates our premise that the task tapped chess-specific abilities of the subjects.

The interaction between fragment type and hemisphere was also significant in the error rates, $F(1,15) = 6.35$, $p < .03$, $r = .55$. Here, contrasts showed that single-chunk fragments were processed more accurately by the right hemisphere (21.4%) than by the left hemisphere (27.6%), $F(1,15) = 40.25$, $p < .001$, $r = .85$, but that there was no significant accuracy difference between hemispheres in processing multiple-chunk fragments (left hemisphere 38.4%, right hemisphere 37.1%), $F(1,15) = 1.69$, $p > .20$, $r = .32$, so there was no speed/accuracy tradeoff. In general, single-chunk fragments were processed more accurately than multiple-chunk fragments, $F(1,15) = 22.57$, $p < .001$, $r = .78$, for the main effect of fragment type. Interestingly, fragments of endgame positions were processed more accurately than were fragments of middlegame positions (25.7% versus 36.7% errors), $F(1,15) = 16.72$, $p < .01$, $r = .73$. It is possible that the more complicated middlegame positions placed greater loads on the subjects' short-term memories, resulting in lower accuracy.

As expected, these results indicate that the right hemisphere is more accurate at parsing chess positions into parts according to a set of acquired chunking rules, and that the left hemisphere is faster at grouping pieces into parts that violate those rules. Despite the strength of this evidence, it is possible that perceptual factors not related to chess knowledge could have accounted for the observed interactions between visual field and fragment type. Perhaps the single-chunk fragments happened to be organized in better accordance with the Gestalt principles than were the multiple-chunk fragments.

An obvious way to address this question would be to test either non-masters on the same task or a group of masters on a variant task with meaningless symbols in place of the chess piece characters in the stimulus diagrams. Unfortunately, both tasks would probably be too difficult for participants, so we performed a regression analysis instead. The dependent variable was the difference between the left- and right-visual-field mean response times across subjects for each of the 64 fragments presented in "yes" trials. These means were computed from the trimmed data set used for the ANOVA discussed above. In addition to type (single- or multiple-chunk), each fragment was coded for seven other independent variables: Number of pieces (2–3), collinearity of pieces (0–2), proximity of pieces (0–1), similarity of piece color (0–1), similarity of piece type (0–2), degree of occupied square color similarity (0–1), and degree of contrast between piece and square colors (0–2). **Figure 15** shows the coding of a sample fragment.

As anticipated, the only variable entered in the stepwise regression was fragment type, $F(1,62) = 4.05, p < .05, R^2 = .06$. A correlation matrix of the variables showed that no other independent variable correlated above $r = 0.11$ with the response time difference. In a similar stepwise regression with error rate difference as the dependent variable, none of the independent variables were entered. These results reinforce the conclusion that the hemispheric asymmetry observed in this experiment is due primarily to the semantic content of the stimuli. They also add support for the hypothesis that the right hemisphere can acquire and apply sets of context-specific default parsing rules while the left hemisphere is superior at overriding such rules when necessary.

The results of Experiments 5 and 6 can be interpreted within the default/override framework for visual parsing proposed by Van Kleeck and Kosslyn (1989). Experiment 5 showed that the left hemisphere was superior to the right hemisphere at recognizing random but *not* normal chess positions. This finding contradicts the hypothesis that the parsing rules used in chess chunking must be used by the left hemisphere to override the visual system's default, right-hemisphere parsing rules (which yield groupings that conform largely to the Gestalt laws of perceptual organization). Experiment 6 showed directly that the right hemisphere was superior at identifying chess position fragments containing a single chunk, but that the left hemisphere was better with multiple chunk fragments of equivalent size. This finding strongly supports the hypothesis that the right hemisphere is more adept at acquiring the chess-specific chunking rules, which the visual system uses in place of its defaults within the context of chess.[25] A consistent pattern across Experiments 5 and 6 emerged as well: The normal positions (Experiment 5) single-chunk fragments (Experiment 6) yielded a right-hemisphere advantage in both response time and error rate measures; whereas the random positions and multiple-chunk fragments both caused opposite (though non-significant) trends in speed and accuracy. It is possible that the less organized stimuli provoke a competition in processing between chess-specific mechanisms and general visual parsing strategies, resulting in poorer and more variable performance.

One could argue that the results of Experiments 5 and 6 show only that the right hemisphere is better than the left at "easy" versions of two perceptual tasks. Indeed, it is

---

[25] For very similar findings in a music task, suggesting that this hemispheric specialization is a principle of perceptual processing in multiple sensory modalities, see Peretz and Babaï (1992).

often suggested that the left hemisphere is specialized for “hard” tasks, but recent experiments involving judgements of different types of spatial relations have provided strong evidence against this idea. For example, Kosslyn et al. (1989, Experiment 3) presented subjects with identical sets of stimuli, a small dot and a horizontal bar, and asked them to decide either whether the dot was above the bar or whether it was within a specified distance from the bar. The metric task (distance judgement) was more difficult, as it engendered greater response times and error rates, but was performed better by the right hemisphere. By contrast, the easier categorical task (above/below judgement) was performed better by the left hemisphere. Therefore, task difficulty alone cannot explain hemispheric differences in perceptual processing unless we consider it a factor in some domains, such as chess, but not in others, such as judging spatial relations.

Neuropsychological findings and the results of two divided visual field experiments have provided converging evidence that the right cerebral hemisphere is critical for chess skill. Chunking, the expert player’s ability to rapidly and efficiently encode positions into memory, is a fundamental component of such skill because representations of positions must be used in other important operations in chess cognition such as position evaluation, move selection, and forward search. Chunking in chess and similar domains can be understood as the imposition of a first-order perceptual organization on the stimulus that arranges its elements into potentially useful groupings. Thus, we conclude that the right hemisphere is critical for chess skill because it is best at using chunking to encode normally-structured positions into memory.

Finally, we offer a speculative explanation of the disproportionately small number of women who play in tournaments and reach the highest levels of achievement in chess.

The surprising fact that fewer than two percent of the world's grandmasters are women, despite chess being not primarily an athletic sport, is often ascribed to social and cultural factors. Our findings on hemispheric specialization for chess perception can be combined with findings on the variability of functional cerebral asymmetry during the menstrual cycle to provide a neuropsychological account for this phenomenon.

Heister, Landis, Regard, and Schroeder-Heister (1989) found that the usual right-hemisphere advantage for a face decision task gradually declined throughout their subjects' menstrual cycles, actually reversing in the premenstrual phase. This change was caused by a decline in right hemisphere performance (not an increase in left hemisphere performance), suggesting that physiological changes during the cycle decrease the "activation" of the right hemisphere and retard its performance on tasks to which it is otherwise suited. Chiarello, McMahon, and Schaefer (1989) obtained a similar pattern of results with a line orientation task, but both groups observed no cyclical change in the usual left-hemisphere superiority on lexical decision tasks. According to our theory of chess skill, the possible deactivation of the right hemisphere could adversely affect the pattern-recognition abilities of female chess players during the later phases of their menstrual cycles. In long competitions such as matches and tournaments, this could result in poorer overall performance. Further development of these ideas awaits replication and greater understanding of the asymmetry shift results, as well as studies of the chess skills of female players. We can be certain, however, that the use of concrete computational theories like the visual parsing model of Van Kleeck and Kosslyn (1989) which was extended in Experiments 5 and 6, is a promising approach.

Functional Neuroanatomy of Chess Expertise

Experiments 5 and 6 strongly suggest that right-hemisphere mechanisms are crucial for the pattern-recognition component of chess skill. They also give indirect evidence for a left-hemisphere role in extracting meaning, at least in cases (random positions, groupings that violate chunk boundaries) when the interpretation required cannot be supported by the pure chunk representation. But beyond hemispheric asymmetry, can we be more precise about the locations of mental cartoons? We should note that our use of a singular noun, “cartoon,” to describe what is likely to be a complex, multipart structure, does not imply that the structure’s parts must all be located in close proximity. Given that the cartoon proper is a spatially-organized arena where perception, imagery, memory, and thought interact, it may be located in both the parietal and frontal lobes, specifically in subregions of each that are directly connected to each other.

The pattern-recognition mechanisms that classifies chess positions and extracts their salient, identifying properties have been frequently analogized to ones presumed to exist for human faces. Right-hemisphere advantages in both domains support this. Recent work suggests that an area in the right temporal lobe, the “fusiform face area,” or FFA, may be specialized for face processing (e.g., Kanwisher, McDermott, & Chun, 1997), but other sorts of stimuli may be able to invoke similar levels of selective activation in this area for subjects trained to discriminate them (e.g., Gauthier et al., in press). For chess masters, chess positions may have this property, so it would be of interest to see whether, on a subject-by-subject basis, the FFA can be activated by chess positions of different sorts in chess masters but not in novices.

The frequently-reported introspective experiences of dynamic, automatic motion by chess players suggest that a motion-processing area such as the middle temporal lobe, which is well-connected to the parietal lobe, may also be involved. Indeed, Kourtzi and Kanwisher (1998) reported an fMRI study showing that more MT activation occurs when subjects view photographs that "imply motion" than when they view photographs that do not (they give the example of athletes in action or at rest). Goebel et al. (1998; see also O'Craven and Kanwisher, 1997) also found more MT activation with fMRI when subjects were explicitly told to imagine static stimuli in motion than when they were not. For skilled chess players, depictions of chess positions would be expected to imply motion of the same sort. Perhaps, since automatically imagined patterns of movement can continue for minutes in the case of chess, MT activation equal to that from a set of implied-motion scenes could be provoked in chess masters by presenting a much smaller set of chess positions.

For example, three types of positions could be created: tactical positions, in which there was a sacrificial variation leading to checkmate; strategic positions, with no such continuation; and positions generated by randomly rearranging the pieces in the positions from the former two types. Following the methodology of Epstein and Kanwisher (1998), each position would be depicted photographically, overlaid with a 10 x 10 grid of black lines, and paired with a control stimulus created by rearranging the elements within the grid so that the stimulus was no longer recognizable as a chess position but still contained the same low-level visual properties. Subjects would view these six types of stimuli for a few seconds per stimulus in alternating epochs, interspersed with periods of fixation, for several averaged runs. MT activation, as measured by the ratio of percent signal increase from fixation between the chess and control stimuli, should be highest for the tactical positions and lowest for the random positions—for chess masters only. Non-players should show no difference among these comparisons. (Eye movements could be recorded in prior offline testing to verify that they were equal across the three conditions.)

To date only two functional neuroimaging studies have been done with chess materials. A PET study by Nichelli et al. (1994) provides confusing but hopeful evidence bearing on the predictions discussed above. Among four conditions compared in a subtractive hierarchy were a “rule retrieval” condition, in which participants shown a chess position decided whether a specified piece could capture another, and a “checkmate” condition, in which they decided whether any move was available that gave immediate checkmate. In the latter case, it is reasonable to assume that several alternative moves must be mentally executed, whereas in the former, only one must be imagined. Both should exercise the chess cartoon, but checkmate should selectively stress it more than rule retrieval, so the subtraction might highlight areas subserving the cartoon representation. In fact, the occipital-parietal junction (areas 7, 18, and 19) and the frontal eye fields were active bilaterally, along with the left orbital-frontal cortex and the right prefrontal cortex. The eye field activation may reflect eye movements associated with increased image scanning in the cartoon or across the visible representation of the chess board, and most of the other areas are consistent with the considerations discussed above. Unfortunately, since the Nichelli et al. study was not designed to test these predictions, it is impossible to draw strong conclusions.

Onofrj et al. (1995) studied five strong amateur chess players (average rating approximately 1975) with SPECT while they examined a chess position and thought about what move they would play over the course of 30 minutes (coincidentally, they used one of the famous positions we used in Experiments 2 and 3; they did not report whether any of their participants were familiar with it). Increased activation beyond a cerebellar baseline was found in the right dorsolateral prefrontal area (11%) and right middle temporal area (5%), with the hemispheric difference reversed for the one left-handed participant. As with the Nichelli et al. study, this rudimentary imaging experiment offers tentative support for the proposed localization of the mental cartoon structure. Area MT may have been activated here but not there because in this study the participants probably repeatedly imagined long sequences of moves and countermoves during the 30 minutes, whereas in Nichelli et al. they only had to test single moves.

The overall findings of significant frontal lobe involvement and almost no left-hemisphere involvement in the neuropsychological studies reviewed here are consistent with the working memory results of Robbins et al. (1996) and the claim that mental cartoons are a type of semi-depictive spatial/abstract representation rather than a verbal code or hybrid spatial/verbal mechanism. Clearly, exploring the functional neuroanatomy of expertise in chess and other domains will be a valuable direction for future research.

Conclusions

The goals of this thesis have been to show that research on chess expertise needs a new theoretical framework, to propose such a framework, to test some of its implications empirically, to investigate how the theory might be implemented in the brain, and to suggest additional experiments that would profitably continue this enterprise. In conclusion, I will try to evaluate the overall success of this effort in the context of other theoretical approaches to expertise.

Throughout, I have argued that the standard theory of chess skill (Simon & Chase, 1973) centered on the concept of chunking would be unable to predict many experimental results that I and other investigators have obtained. However, it must be noted that this theory was never designed to predict some of them. In particular, it has little to say about specific processes of visual imagery in chess, and nothing to say directly about neural mechanisms. So in these areas it must be regarded as incomplete rather than incorrect. It also deserves praise for its parsimony in attempting to stretch a single mechanism as far as possible to account for expertise effects perception and memory organization. In my view though, it ultimately fails even in this area, and a division of expert memory into separate pattern and conceptual stores provides a superior account.

Progress by fractionation is a common trend in cognitive neuroscience; the decomposition of high-level visual processing into two major systems (e.g., Ungerleider & Mishkin, 1982; Milner & Goodale, 1995), or working memory into multiple subsystems (Baddeley, 1986), were important theoretical advances brought about by the inability of simpler theories to account for new experimental findings. However, the chunking theory is not replaced entirely by my account; it simply becomes a component of it, in much the same way as the classic principles of short-term memory were incorporated into working memory theory rather than discarded entirely.

Although the chunking theory is the most influential and accepted theory of expertise in chess, several others have been proposed. Gobet (1998b) summarizes and reviews “SEEK” (Holding, 1985), long-term working memory (Ericsson & Kintsch, 1995), and his own updated version of chunking, the “template theory” (Gobet & Simon, 1996d); to his list I would add de Groot’s original theory (1946), as well as the “constraint attunement hypothesis” proposed by Vicente and Wang (1998). I have not dealt with these alternatives in detail thus far because in general, they apply only to expert performance in memory tasks, making no attempt to incorporate visualization, thinking ahead, or neural mechanisms. Moreover, Gobet (1998b; see also Gobet & Simon, 1998a; Simon & Gobet, 1999) argues persuasively that the template theory is superior to all of these alternatives in accounting for memory effects.

The template theory essentially adds to the chunking theory a new data structure similar to an artificial intelligence-style “frame”—a schema with fixed elements and variable slots that can be filled during memory encoding. These templates in a chess master’s memory correspond to familiar categories of positions from standard openings, and are invoked to explain how very skilled chess players can seemingly take in information at a much faster rate than the original chunking theory predicted. (The style of computer simulation originated by Herbert Simon and Allen Newell emphasized modeling the actual time consumed for cognitive operations.) This is perfectly fine as a model for one type of conceptual knowledge that must coexist with chunks, and its addition to the classical model is an implicit admission that chunking theory is incomplete in at least one way. However, it is based on the EPAM discrimination network, which is not the most biologically plausible mechanism for implementing pattern recognition in the brain. It also does not account for any forms of conceptual knowledge other than templates, which are limited in number and specific to classes of opening positions. Such templates could not explain the results of Experiment 2, for example, because the relevant changes made to the famous positions did not displace them from one opening class into another, yet they were noticed more than the irrelevant changes. Still, the challenge posed by the template theory exposes the main weakness of the theory proposed here: it is not yet specified in enough detail to permit computer implementation. Along with applying the theoretical ideas to other domains besides chess, implementing them must be a primary goal of future work.

The mental cartoons hypothesis does have several advantages over previous theories. It was formulated in part by considering the introspective experiences of experts, which directly inspired its main features. More importantly, it deliberately encompasses a wider variety of differences in cognition between experts and novices. Although it is no longer believed that people with extreme skill in a limited domain are somehow abnormal (except in rare cases of savantism or perhaps Asperger's syndrome), it is still thought that expertise comes only from having more, better, or more efficiently organized knowledge about a domain. From this flows the voguish theory that deliberate practice is the only important variable in determining how skilled an individual will become (e.g., Ericsson & Charness, 1994). The evidence presented here, especially the finding that chess masters are superior to novices in their ability to visualize actions within their domain of expertise (Experiment 1), suggests otherwise. It may not be unreasonable to conclude by suggesting the “expertise equals cognition hypothesis,” that *expertise is nothing more or less than the application and adaptation of the full range of relevant cognitive abilities to the constraints and demands an unusual problem domain.*

Expertise is a pervasive psychological phenomenon, so much so that we can easily forget how central it is to our concept of self and our habits of social interaction. The first question you ask of someone you meet is “what do you do?” The theory and evidence presented here represent the first step towards a new answer to another common question about experts, “how do they do it?”

Acknowledgements

Experiment 1 was conducted in collaboration with Bassina Farbenblum and Susan Curry; it was supported by grants to Stephen M. Kosslyn. Experiments 2 and 3 were conducted in collaboration with Daniel J. Benjamin and Daniel J. Simons; they were supported by the Harvard University Ditmars Bequest. Experiment 4 was conducted in collaboration with Eliot Hearst and Mark E. Glickman; it was supported by the Harvard University McMasters Fund and Psychophysics Fund. Experiments 5 and 6 were conducted in collaboration with Sania E. Hamilton; they were supported by grants to Stephen M. Kosslyn. Most of the description of Experiment 6 was published, in somewhat different form, in *Neuropsychologia* (Chabris & Hamilton, 1992).

Stephen M. Kosslyn and Sheldon H. White provided valuable support, advice, and assistance throughout the work described here, and gave very useful comments on earlier versions of this manuscript. Daniel J. Simons, Ken Nakayama, Rich McNally, Kenneth Norman, Patrick Wolff, Jonathan Yedidia, Eliot Hearst, Mark Glickman, Jennifer Shephard, Annya Hernandez, Hon-Ming Eng, Giorgio Ganis, Steve Mitroff, Phyra McCandless, Stuart Margulies, Lee Cranberg, and especially Daniel J. Benjamin have also been extremely helpful. The term "mental cartoon" occurred to me after discussions with David Mumford and Stephen M. Kosslyn regarding mathematical expertise. I have benefited from many other conversations with chess players, mathematicians, computer scientists, and psychology researchers over the years since I began thinking about the issues addressed in this thesis. Finally, I gratefully acknowledge the support of an NSF Graduate Fellowship, and most of all my parents and friends.

References

[References marked with an asterisk (*) served as sources for the famous chess positions used in Experiments 2 and 3, but may not be cited directly in the text.]

Abrahams, G. (1952). *The chess mind*. New York: Pellegrini & Cudahy.

*Alekhine, A. (1962). *The book of the Nottingham international chess tournament.* New York: Dover.

Anderson, J.R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85(4)*, 249–277.

Anzai, Y. (1991). Learning and use of representations for physics expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 64–92). Cambridge: Cambridge University Press.

*Averbakh, Y. (1992). *Chess tactics for advanced players.* Dallas: Chess Digest.

Bachmann, T., & Oit, M. (1992). Stroop-like interference in chess players' imagery: An unexplored possibility to be revealed by the adapted moving-spot task. *Psychological Research*, *54(1)*, 27–31.

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

Beaumont, J.G. (Ed.) (1982). *Divided visual field studies of cerebral organization*. New York: Academic Press.

Berliner, H.J. (1993). Playing computer chess in the human style. *ICCA Journal, 16(3),* 176–182.

Bergson, H. (1902). L'effort intellectuel. [Mental effort.] *Revue Philosophique de la France et de L'Etranger, 13,* 1–27.

Binet, A. (1893). Les grandes mémoires: Résumé d'une enquête sur les joueurs d'echecs. *Revue des Deux Mondes*, *117*, 826–859. (Translated by M.L. Simmel and S.B. Barron as "Mnemonic virtuosity: A study of chess players," *Genetic Psychology Monographs*, *74,* 127–162, 1966.)

Binet, A. (1894). *Psychologie des grands calculateurs et joueurs d'échecs.* [Psychology of great calculators and chess players.] Paris: Hachette.

Blumenfeld, B. (1998). Visual imagination and chess analysis. (Translated by J. Sugden from the Russian version first published in 1936.) In M. Dvorestky & A. Yusupov, *Attack and defence: How creative thought develops in a chess player* (pp. 46–52). London: Batsford.

*Botvinnik, M.M. (1984). *Half a century of chess.* Oxford: Pergamon.

Botvinnink, M.M. (1993). Three positions. *ICCA Journal, 16(2),* 71–75.

Bradley, R., & Terry, M. (1952). The rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika, 39,* 324–345.

Bransford, J.D., & Johnson, M.K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726.

*Bronstein, D. (1979). *Zurich international chess tournament 1953*. New York: Dover. (Translated by J. Marfia from the second Russion edition published in 1960.)

Bronstein, D. (1993). Mimicking human oversight. *ICCA Journal, 16(3),* 183–184.

Bronstein, D., & Smolyan, G. (1982). *Chess in the eighties*. Oxford: Pergamon.

Cabanne, P. (1971). *Dialogues with Marcel Duchamp.* London: Thames and Hudson.

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

Campbell, M.S. (1988). Chunking as an abstraction mechanism. Technical Report CMU-CS-86-116, Computer Science Department, Carnegie Mellon University.

Capablanca, J.R. (1938). On chess and chess players. *Chess*, 14 March, 235.

Chabris, C.F. (1994a). Perception and memory of chess positions: Preliminary reinvestigations. Unpublished manuscript.

Chabris, C.F. (1994b). The early history of chess psychology. Unpublished manuscript.

Chabris, C.F. (1999). Revenge of the words: Introspective reports as data in cognitive neuroscience. Manuscript in preparation.

Chabris, C.F., & Hamilton, S.E. (1992). Hemispheric specialization for skilled perceptual organization by chessmasters. *Neuropsychologia*, *30*, 47–57.

Chabris, C.F., & McCandless, P.M. (1999). [Recognition memory in chess masters and novices for photographs and diagrams of chess positions.] Unpublished raw data.

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

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, *7(2)*, 467–476.

Charness, N. (1992). The impact of chess research on cognitive science. *Psychological Research*, *54(1)*, 4–9.

Charness, N., & Schultetus, R.S. (1998, November). Life-span development of chess skill: Aging affects the more skilled sooner. Presented at the Annual Meeting of the Psychonomic Society, Dallas, TX.

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.

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.

Chiarello, C., McMahon, M.A., and Schaefer, K. (1989). Visual cerebral lateralization over phases of the menstrual cycle: A preliminary investigation. *Brain and Cognition, 11,* 18–36.

Church, R.M., & Church, K.W. (1983). Plans, goals, and search strategies for the selection of a move in chess. In P.W. Frey (Ed.), *Chess skill in man and machine* (2nd ed.) (pp. 131–156). New York: Springer.

Condon, J.H., & Thompson, K. (1983). Belle. In P.W. Frey (Ed.), *Chess skill in man and machine* (2nd ed.) (pp. 82–118). New York: Springer.

Cooke, N., 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(3)*, 321–351.

Costin, D. (1988). MacLab: A Macintosh system for psychology labs. *Behavior Research Methods, Instruments, & Computers, 20,* 197–200.

Cranberg, L.D., & Albert, M.L. (1988). The chess mind. In L.K. Obler & D. Fein (Eds.), *The exceptional brain* (pp. 156–190). New York: Guilford.

Dallenbach, K.M. (1917). Blindfold chess: The single game. In *Studies in psychology contributed by colleagues and former students of Edward Bradford Titchener* (pp. 214–230). Worcester, MA: Louis N. Wilson.

de Groot, A.D. (1938). Begaafdheid voor het schaakspel. [The gift for chess.] *Algemeen Nederlands Tijdschrift voor Wijsbegeerte en Psychologie, 32(2),* 97–105.

de Groot, A.D. (1946). *Het denken van de schaker.* [The thought of the chess player.] Amsterdam: North-Holland. (Updated translation published as *Thought and choice in chess*, Mouton, The Hague, 1965; corrected second edition published in 1978.)

de Groot, A.D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory* (pp. 19–50). New York: Wiley.

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

*Donaldson, J., & Minev, N. (1994). *Akiba Rubinstein: Uncrowned king.* Seattle: International Chess Enterprises.

Dror, I.E. (1992). Visual mental rotation: Different processes used by pilots. *Proceedings of the 36th Annual Meeting of the Human Factors Society* (vol. 2) (pp. 1368–1372).

Dror, I.E., Kosslyn, S.M., & Waag, W.L. (1993). Visual-spatial abilities of pilots. *Journal of Applied Psychology, 78(5),* 763–773.

Egan, D.E., & Schwartz, B.J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition, 7,* 149–158.

Eglen, J. (1993). [Performance of skilled chess players on the WAIS-R.] Unpublished raw data.

Ellis, R.R., & Lederman, S.J. (1998). The golf-ball illusion: Evidence for top-down processing in weight perception. *Perception, 27(2),* 193–201.

Ellis, S.H. (1973). Structure and experience in the matching and reproduction of chess patterns. *Dissertation Abstracts, 23,* 2335-B.

Elo, A.E. (1961, June). New USCF rating system. *Chess Life,* 160–161.

Elo, A.E. (1986). *The rating of chessplayers, past and present* (2nd ed.). New York: Arco.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598-601.

Ericsson, K.A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49(8)*, 725–747.

Ericsson, K.A., & Delaney, P.F. (1998). Working memory and expert performance. In R.H. Logie & K.J. Gilhooly (Eds.), *Working memory and thinking* (pp. 93–114). Hove, UK: Psychology Press.

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

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

Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102(2)*, 211–245.

Ericsson, K.A., Krampe, R.T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100(3)*, 363–406.

*Euwe, M. (1976). *From Steinitz to Fischer.* Belgrade: Chess Informant.

*Euwe, M., & Nunn, J. (1997). *The development of chess style.* London: Batsford.

*Evans, L., & Smith, K. (1973). *Chess world championship 1972: Fischer versus Spassky.* New York: Simon & Schuster.

Fine, R. (1965). The psychology of blindfold chess: An introspective account. *Acta Psychologica*, *24*, 352–370.

Finke, R.A., & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 142–147.

Finke, R.A., & Pinker, S. (1983). Directional scanning of remembered visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 398–410.

Freyhof, H., Gruber, H., & Ziegler, A. (1992). Expertise and hierarchical knowledge representation in chess. *Psychological Research*, *54(1)*, 32–37.

Frydman, M., & Lynn, R. (1992). The general intelligence and spatial abilities of gifted young Belgian chess players. *British Journal of Psychology*, *83*, 233–235.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., & Gore, J.C. (in press). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience.*

Gazzaniga, M.S. (1985). *The social brain.* New York: Basic Books.

Ginsburg, M. (1995). Grandmaster chess recall under extreme conditions: Evidence for conceptual categorization and implications for the chunking theory of memory. Unpublished manuscript.

Glickman, M.E. (1995). Chess rating systems. *American Chess Journal, 3,* 59–102.

Glickman, M.E., & Jones, A.C. (in press). Rating the chess rating system. *Chance.*

Gobet, F. (1998a). Chess players' thinking revisited. *Swiss Journal of Psychology, 57(1),* 18–32.

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

Gobet, F. (in press). Marriage of two passions: The chess, psychology, and chess psychology of Adrian de Groot. *American Chess Journal.*

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.H.M. Uiterwijk (Eds.), *Advances in Computer Chess 7* (pp. 35–60). Maastricht, The Netherlands: University of Limburg.

Gobet, F., & Simon, H.A. (1996a). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review, 3(2),* 159–163.

Gobet, F., & Simon, H.A. (1996b). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grand-master-level chess. *Psychological Science, 7(1),* 52–55.

Gobet, F., & Simon, H.A. (1996c). Recall of random and distorted chess positions: Implications for the theory of expertise. *Memory & Cognition, 24(4),* 493–503.

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

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

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., & Singer, W. (1998). The constructive nature of vision: Direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal of Neuroscience, 10(5),* 1563–1573.

Goldin, S.E. (1979). Recognition memory for chess positions: Some preliminary research. *American Journal of Psychology, 92,* 19–31.

*Golombek, H. (1996). *Capablanca's best games.* London: Batsford.

*Goodman, D., & Keene, R. (1997). *Man versus machine: Kasparov versus Deep Blue.* Cambridge, MA: H3 Publications.

*Gordon, S.W. (1997). *Samuel Reshevsky.* Jefferson, NC: McFarland.

Hartmann, D. (1989). Notions of evaluation functions tested against grandmaster games. In D.F. Beal (Ed.), *Advances in Computer Chess 5* (pp. 91–141) Amsterdam: North-Holland.

Hartston, W.R., & Wason, P.C. (1985). *The psychology of chess.* New York: Facts on File.

Hecht, H., & Proffitt, D.R. (1995). The price of expertise: Effects of experience on the water-level task. *Psychological Science, 6(2),* 90–95.

Heinz, E.A. (1998). DarkThought goes deep. *ICCA Journal, 21(4),* 228–244.

Heister, G., Landis, T., Regard, M., & Schroeder-Heister, P. (1989). Shift of functional cerebral asymmetry during the menstrual cycle. *Neuropsychologia, 27,* 871–880.

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

Holding, D.H. (1989). Evaluation factors in human tree search. *American Journal of Psychology*, *102*, 103–108.

Holding, D.H. (1992). Theories of chess skill. *Psychological Research*, *54(1)*, 10–16.

Holding, D.H., & Reynolds, R.I. (1982). Recall or evaluation of chess positions as determinants of chess skill. *Memory & Cognition, 10,* 237–242.

Horgan, D.D., & Morgan, D. (1990). Chess expertise in children. *Applied Cognitive Psychology, 4,* 109–128.

Hsu, F-H., Anantharaman, T., Campbell, M., & Nowatzyk, A. (1990, October). A grandmaster chess machine. *Scientific American, 263,* 44–50.

Hyatt, R.M., & Newborn, M. (1997). Crafty goes deep. *ICCA Journal, 20(2),* 79–86.

Kämpf, U., & Strobel, R. (1998). “Automatic” position evaluation in “controlled” change detection: Data-driven vs. concept-guided encoding and retrieval strategy components in chess players with varying degrees of expertise. *Zeitschrift für Psychologie, 206(1),* 23–46.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17(11),* 4302–4311.

*Keene, R., & Goodman, D. (1985). *Maneuvers in Moscow: Karpov–Kasparov II.* New York: Macmillan.

Keres, P. (1991). *Power chess: Great grandmaster battles from Russia.* New York: McKay.

Korb, K.B. (1994). Stephen J. Gould on intelligence. *Cognition*, *52(2)*, 111–123.

Kosslyn, S.M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics, 14(1),* 90–94.

Kosslyn, S.M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

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

Kosslyn, S.M., Ball, T.M., & Reiser, B.J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 47–60.

Kosslyn, S.M., Koenig, O., Barrett, A., Cave, C.B., Tang, J., and Gabrieli, J.D.E. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 723–735.

Kourtzi, Z., & Kanwisher, N. (1998, November). Activation in human MT/MST for static images with implied motion. Presented at the Annual Meeting of the Psychonomic Society, Dallas, TX.

Kremer, M. (1989). The chess career of Marcel Duchamp. *New In Chess, 2,* 34–42.

Krogius, N. (1976). *Psychology in chess*. Great Neck, NY: RHM Press.

Lasker, E. (1932). *Lasker’s chess manual.* London: Printing-Craft.

Lories, G. (1987). Recall of random and non random chess positions in strong and weak chess players. *Psychologica Belgica*, *27-2*, 153–159.

Malkin, V. (1982). How to "see" a chess position. *64.* [In Russian.]

Marks, D.F. (1972). Individual differences in the vividness of visual imagery and their effect on function. In W.P. Sheehan (Ed.), *The function and nature of imagery* (pp. 83–108). New York: Academic Press.

Marks, D.F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64(1),* 17–24.

Marr, D. (1982). *Vision*. New York: Freeman.

Maslow, A.H. (1971). *The farther reaches of human nature*. New York: Viking.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 86,* 287–330.

Mieses, J. (1940). Psychology and the art of chess. *Chess*, April, 154–156.

Milner, B. (1968). Visual recognition and recall after right temporal-lobe excision in man. *Neuropsychologia, 6,* 191–209.

Milner, A.D., & Goodale, M.A. (1995). *The visual brain in action.* Oxford: Oxford University Press.

Milojkovic, J.D. (1982). Chess imagery in novice and master. *Journal of Mental Imagery*, *6*, 125–144.

Mosteller, F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika, 16,* 3–9.

Newborn, M. (1985). A hypothesis concerning the strength of chess programs. *ICCA Journal, 8(4),* 209–215.

Newell, A., Shaw, C., & Simon, H. (1958). Chess playing programs and the problem of complexity. *IBM Journal of Research and Development*, *2*, 320–335.

Nichelli, P., Grafman, J., Pietrini, P., Alway, D., Carton, J.C., & Miletich, R. (1994). Brain activity in chess playing. *Nature*, *369*, 191.

O'Craven, K.M., & Kanwisher, N. (1997). Visual imagery of moving stimuli activates area MT/MST. *Abstracts of the Society for Neuroscience, 23,* 299.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9,* 97–113.

Onofrj, M., Curatola, L., Valentini, G., Antonelli, M., Thomas, A., & Fulgente, T. (1995). Non-dominant dorsal-prefrontal activation during chess problem solution evidenced by single photon emission computerized tomography (SPECT). *Neuroscience Letters, 198,* 169–172.

Peretz, I., & Babaï, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia*, *30*, 277–292.

Pinker, S., Choate, P.A., & Finke, R.A. (1984). Mental extrapolation in patterns constructed from memory. *Memory & Cognition, 12(3),* 207–218.

Reitman, J. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology, 4,* 194–206.

Reynolds, R.I. (1982). Search heuristics of chess players of different calibers. *American Journal of Psychology*, *95(3)*, 383–392.

Reynolds, R.I. (1991). The application of a search heuristic by skilled problem solvers. *Bulletin of the Psychonomic Society*, *29(1)*, 55–56.

Rhodes, G. (1996). *Superportraits: Caricatures and recognition.* Hove, UK: Psychology Press.

Robbins, E. (1994). *Why architects draw.* Cambridge, MA: MIT Press.

Robbins, T.W., Anderson, E.J., Barker, D.R., Bradley, A.C., Fearneyhough, C., Henson, R., Hudson, S.R., & Baddeley, A.D. (1996). Working memory in chess. *Memory & Cognition, 24(1),* 83–93.

Rumelhart, D.E., & McClelland, J.L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Saariluoma, P. (1984). Coding problem spaces in chess: A psychological study. *Commentationes Scientiarum Socialium, 23*, 1–121.

Saariluoma, P. (1992). Visuospatial and articulatory interference in chess players' information intake. *Applied Cognitive Psychology*, *6*, 77–89.

Sacks, O.W. (1985). *The man who mistook his wife for a hat and other clinical tales.* New York: Summit.

Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, *82*, 454–465.

Simon, H.A., & Chase, W.G. (1973). Skill in chess. *American Scientist*, *61,* 394–403.

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

Simon, H.A., & Gobet, F. (1999). Expertise effects in memory recall: A reply to Vicente and Wang. Manuscript submitted for publication.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74.*

Steiner, G. (1974). *Fields of force: Fischer & Spassky at Reykjavik*. New York: Viking.

Taine, H-A. (1875). *On intelligence* (Vol. I). New York: Henry Holt. ("Translated from the French by T.D. Haye and revised with additions by the author." First published as *De l'intelligence*, Hachette, Paris, 1870. Reprinted by University Publications of America, Washington, 1977.)

Thompson, K. (1982). Computer chess strength. In M.R.B. Clarke (Ed.), *Advances in computer chess 3* (pp. 55–56). Oxford: Pergamon.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–286.

Tisdall, J. (1997). *Improve your chess now.* London: Cadogan.

Ungerleider, L.G., & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, & R.J.W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van Kleeck, M.H. (1989). Hemispheric differences in global versus local processing of hierarchical visual stimuli by normal subjects: New data and a meta-analysis of previous studies. *Neuropsychologia, 27,* 1165–1178.

Van Kleeck, M.H., & Kosslyn, S.M. (1989). Gestalt laws of perceptual organization in an embedded figures task: Evidence for hemispheric specialization. *Neuropsychologia, 27,* 1179–1186.

Vasta, R., Rosenberg, D., Knott, J.A., & Gaze, C.E. (1997). Experience and the water-level task revisited: Does expertise extract a price? *Psychological Science, 8(4),* 336–339.

Vicente, K.J., & Brewer, W.F. (1993). Reconstructive remembering of the scientific literature. *Cognition*, *46(2)*, 101–128.

Vicente, K.J., & de Groot, A.D. (1990). The memory recall paradigm: Straightening out the historical record. *American Psychologist*, *45,* 285–287.

Vicente, K.J., & Wang, J.H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review, 105(1),* 33–57.

Walsh, V., & Rushworth, M. (1999). A primer of magnetic stimulation as a tool for neuropsychology. *Neuropsychologia, 37(2),* 125–135.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W.D. Ellis (Ed.), *A source book of Gestalt psychology.* London: Kegan, Paul, Trench, Trubner & Co.

Yusupov, A. (1998). Missed brilliancy prizes. (Translated by J. Sugden from the Russian version.) In M. Dvorestky & A. Yusupov, *Attack and defence: How creative thought develops in a chess player* (pp. 151–175). London: Batsford.

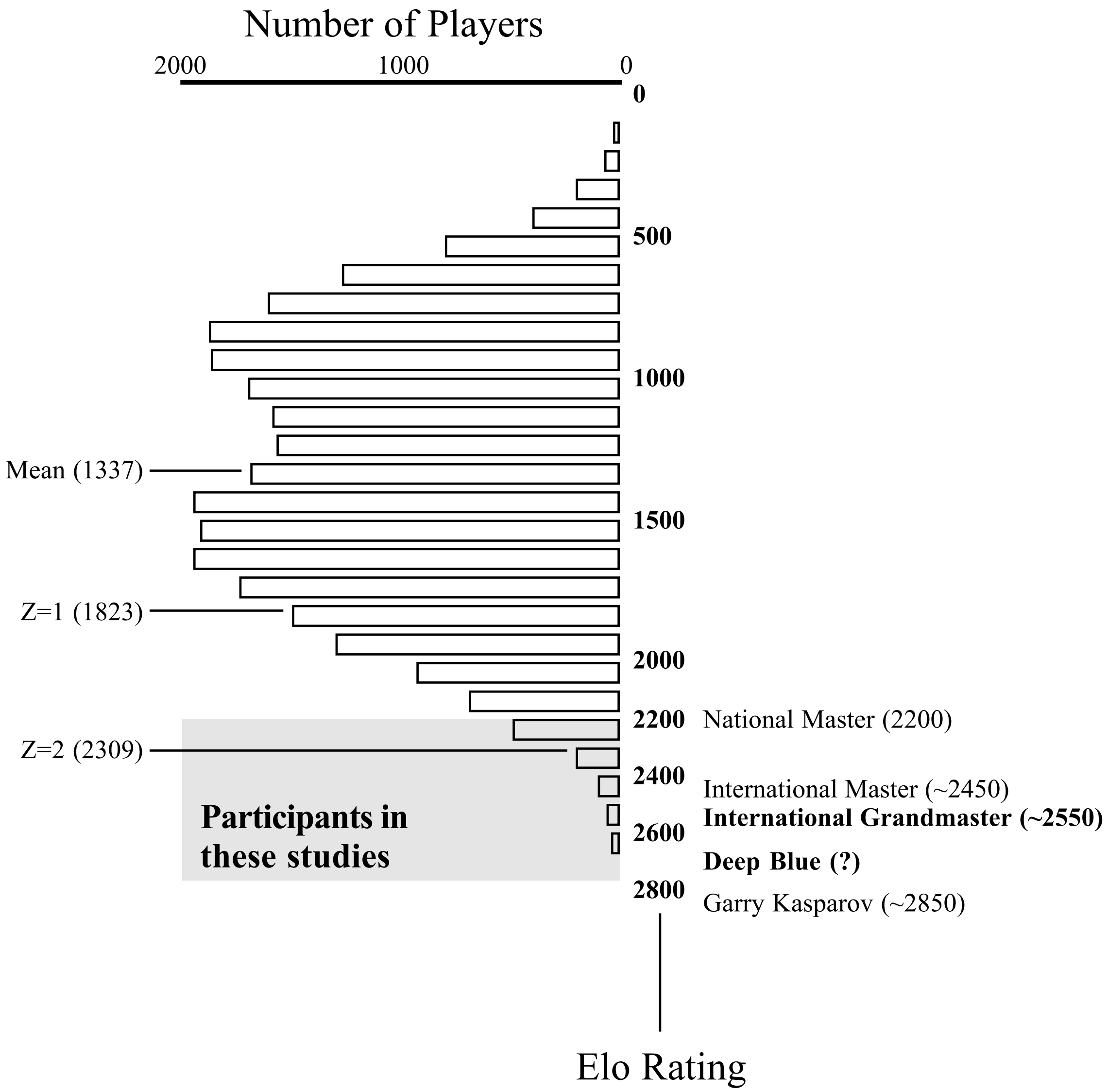


Figure 1. Distribution of chess skill according to the scale developed by Elo (1961, 1986), based on data from the July 1998 U.S. Chess Federation rating list. All participants referred to as “experts” or “masters” in Experiments 1–6 ranged in playing strength from National Master to International Grandmaster and had ratings on this scale of at least 2149. Most of the chess-specific tasks used here could not be completed successfully by chess “novices,” except for Experiment 1, whose “novice” participants had never played in organized chess tournaments and did not posses chess ratings.

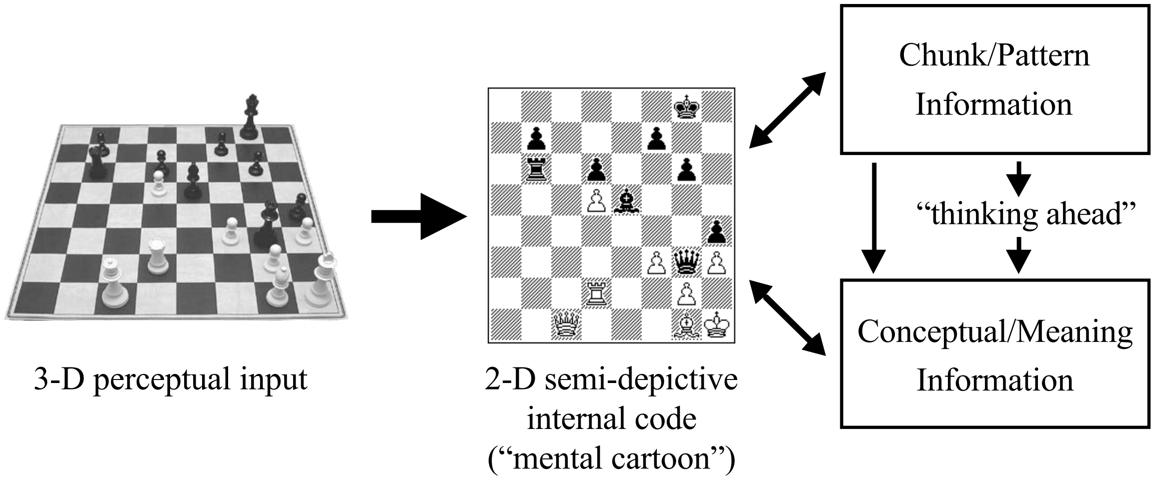


Figure 2. Proposed model of key mechanisms in chess expertise. The theory offered here posits an intermediate two-dimensional diagram representation, or “mental cartoon,” and a separate store of conceptual knowledge, neither of which are included in other theories that have been applied to chess expertise.

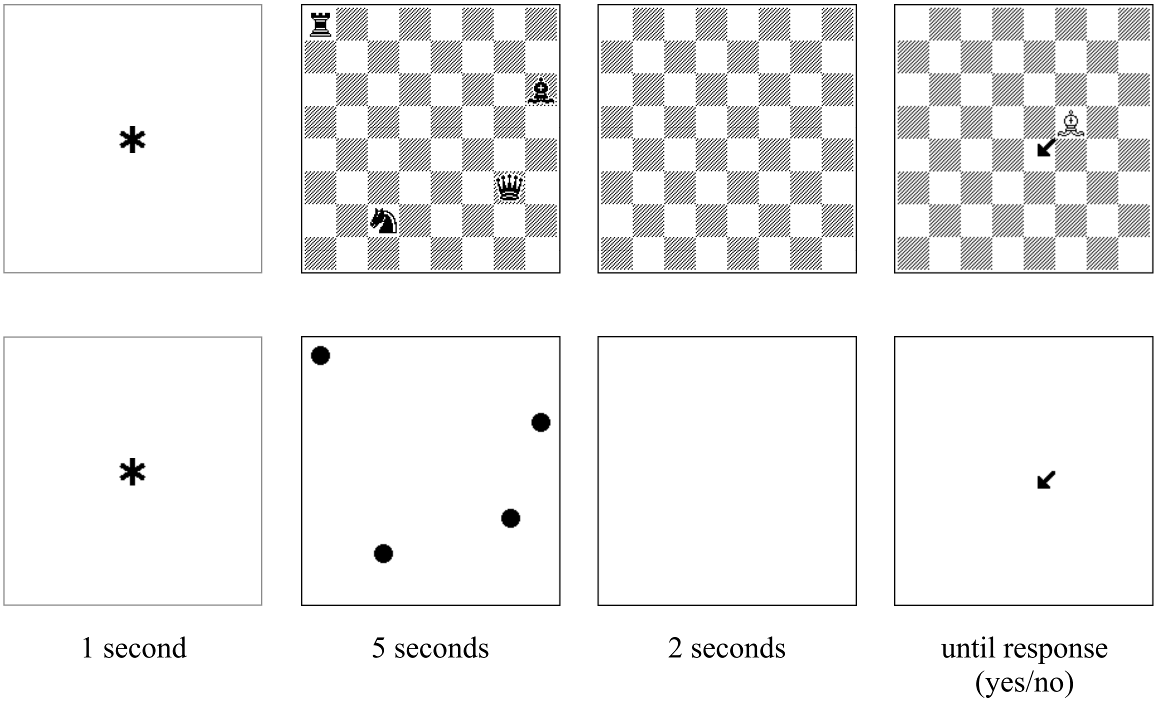


Figure 3. Stimuli and event sequence for trials in Experiment 1. *Top,* Chess scanning task. *Bottom,* Dot scanning task. *Note:* During the fixation point event the box was not actually displayed on the screen.

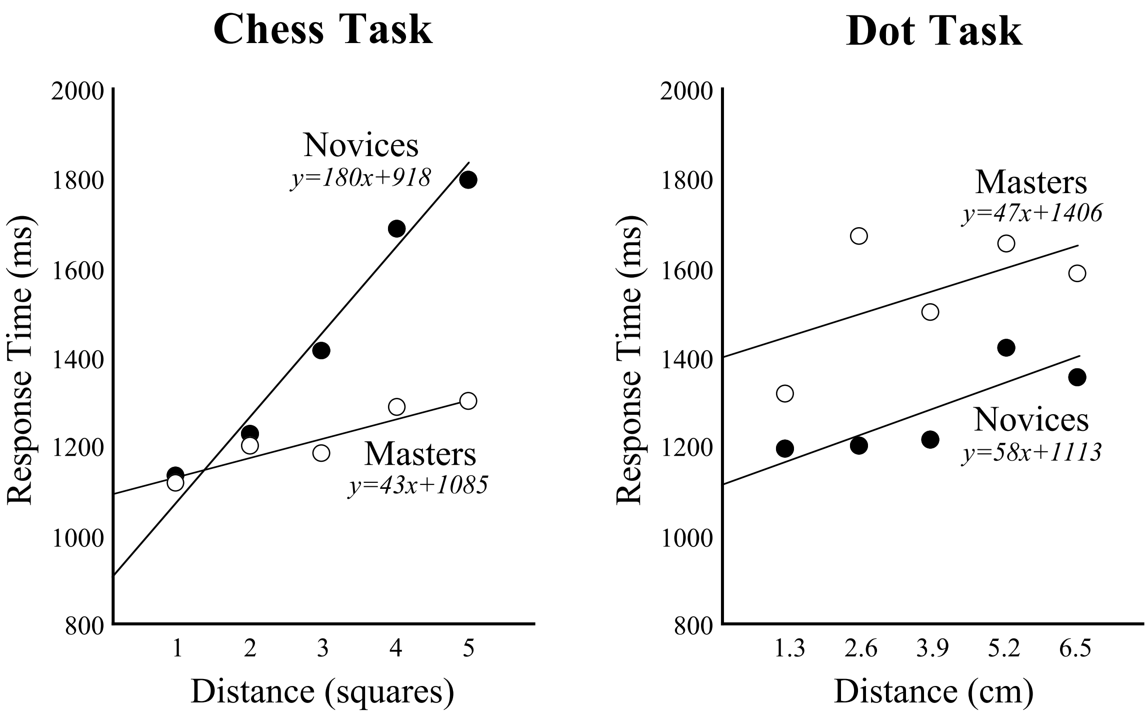
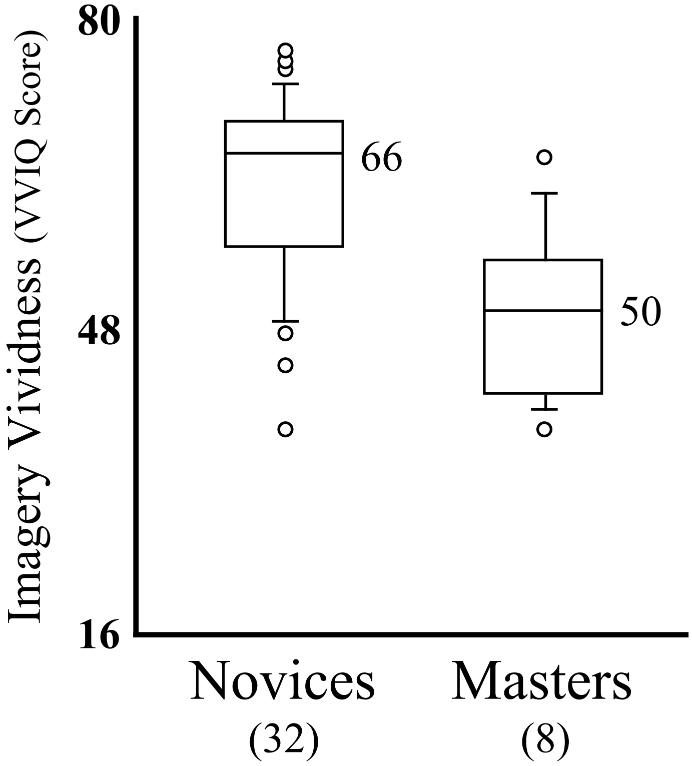


Figure 4. Results of Experiment 1: mean response times of chess masters (6) and novices (18) as a function of distance scanned. *Left,* Chess scanning task. *Right,* Dot scanning task. Regression lines illustrate slope and intercept of scanning functions for each group on each task. Note that for each regression equation, time is a function of distance as measured in chessboard squares; since these were not actually present in the stimuli used for the Dot task, the horizontal axis is labelled with the equivalent distances in centimeters.



Supplemental Figure A. Imagery vividness scores for chess novices and masters. See Experiment 1, footnote 15, p. 46 for more information.

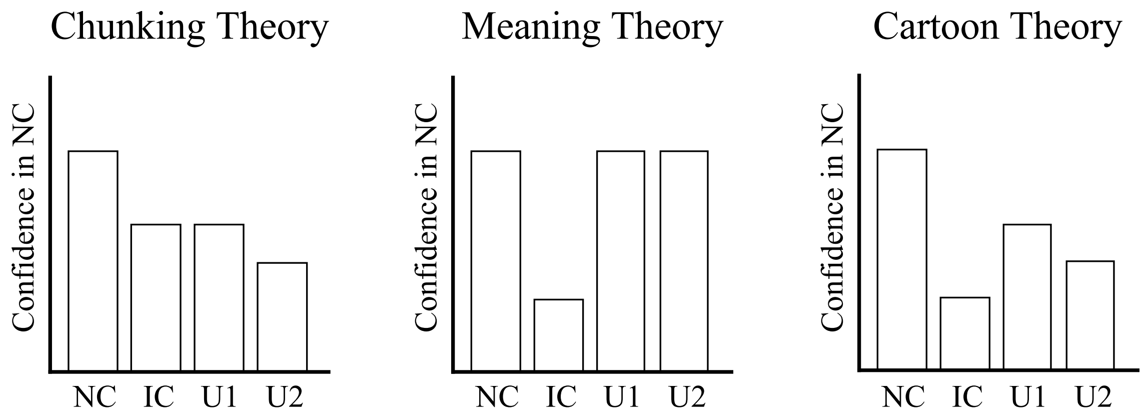
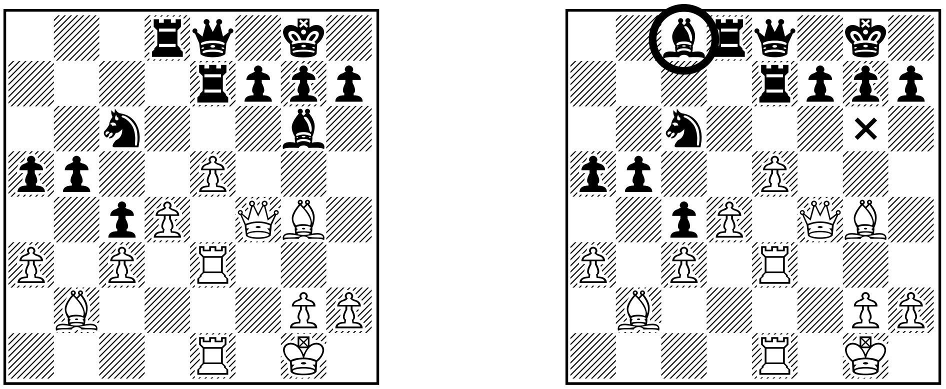
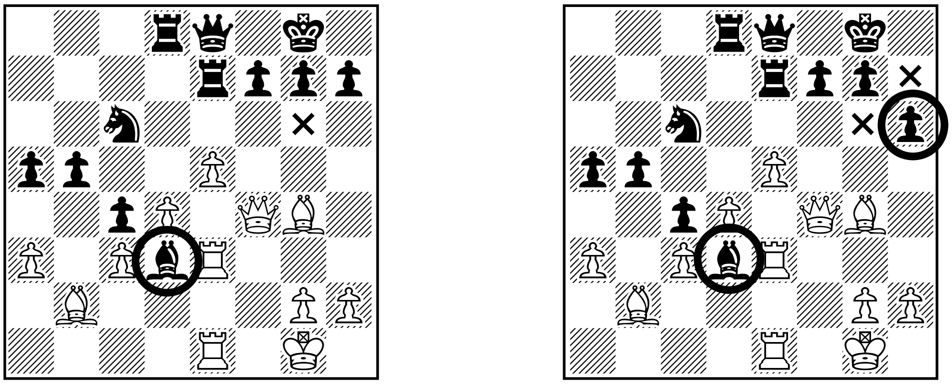


Figure 5. Predicted results for Experiment 2, based on the Chunking, Meaning, and Mental Cartoon theories described in the text, showing confidence that a famous chess position has not been changed as a function of the types of change actually made. *NC,* no change. *RC,* relevant change. *I1,* one irrelevant change. *I2,* two irrelevant changes.



No change (15 trials)

One Relevant change (5 trials)



One Irrelevant change (5 trials)

Two Irrelevant changes (5 trials)

Reshevsky–Petrosian, Zurich 1953, Black to play

Figure 6. Sample stimuli for Experiment 2. The circles and Xs indicate changes that were applied to make the three changed versions of the original famous chess position. The caption shown below the diagrams here was presented to the right of the diagram in the booklet that participants filled out. (See text for explanation.)

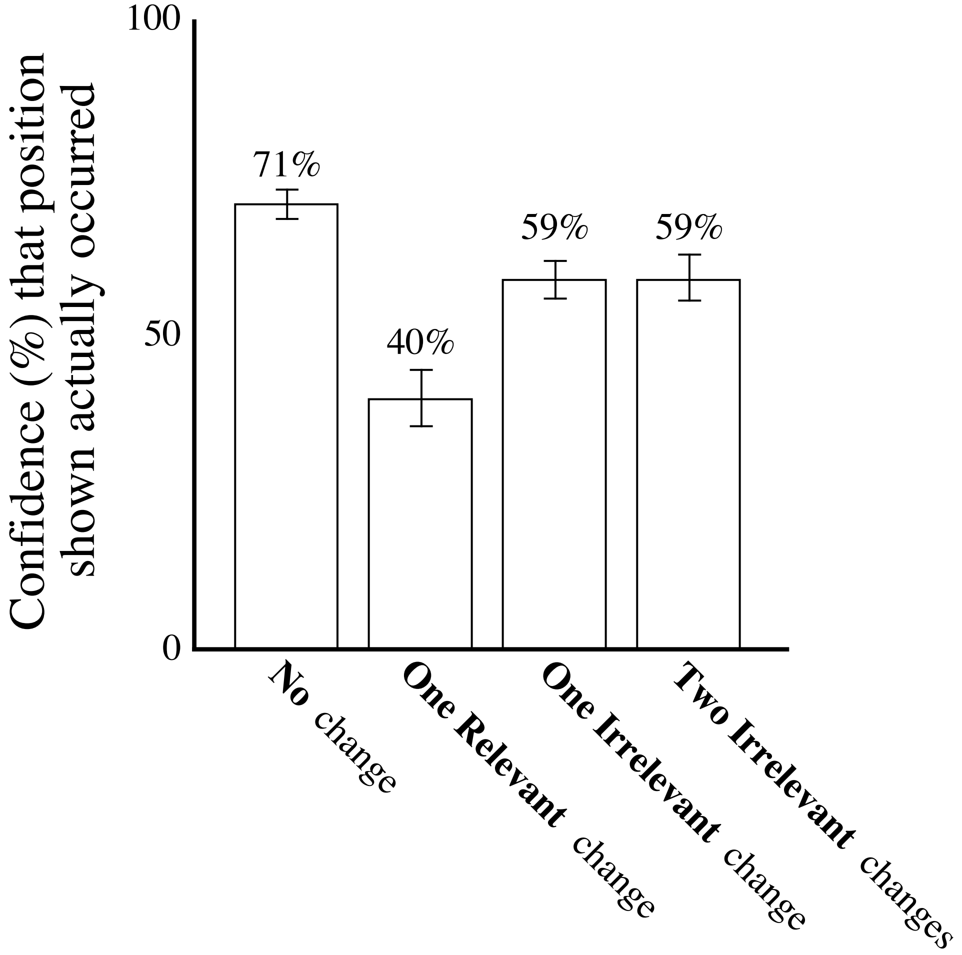


Figure 7. Results of Experiment 2, showing the confidence of 24 chess masters that famous chess positions were shown to them unchanged, as a function of the types of change actually made. Error bars represent standard errors of the mean. (See text for 95% confidence intervals.)

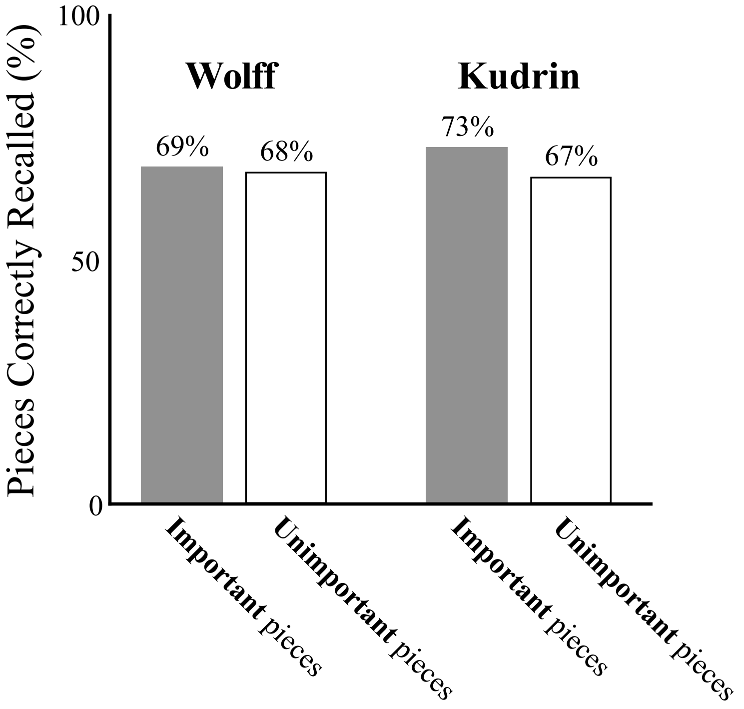


Figure 8. Proportion of important and unimportant pieces recalled correctly (both piece identity and square location correct) from famous chess positions by grandmasters Wolff and Kudrin in Experiment 3.

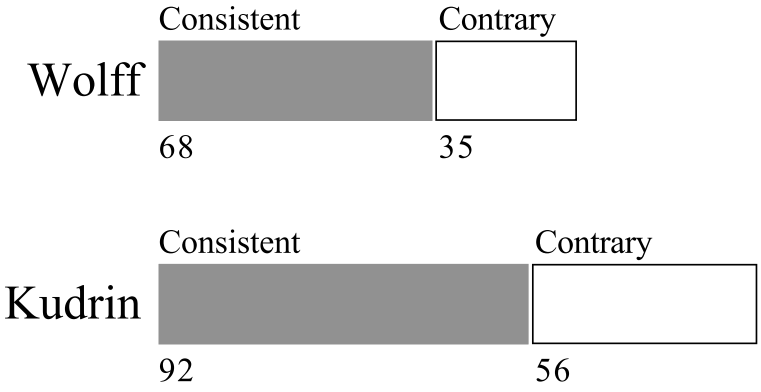
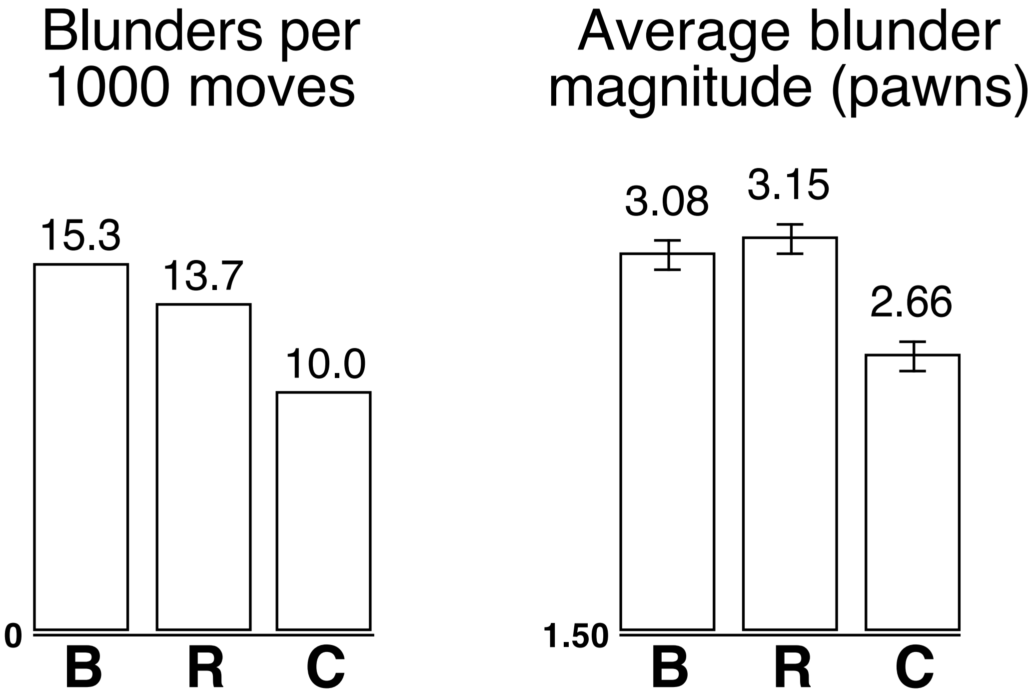


Figure 9. Number of consistent and contrary errors in recall of famous chess positions by grandmasters Wolff and Kudrin in Experiment 3. (See text for explanation.)



Supplemental Figure B. Results of Experiment 4. *B:* Blindfold chess. *R:* Rapid chess. *C:* Classical chess. See Table 3, p. 70 for details.

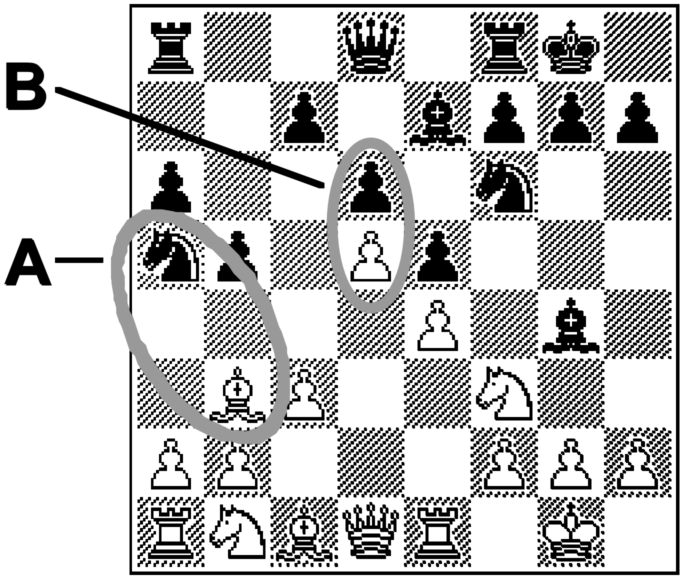
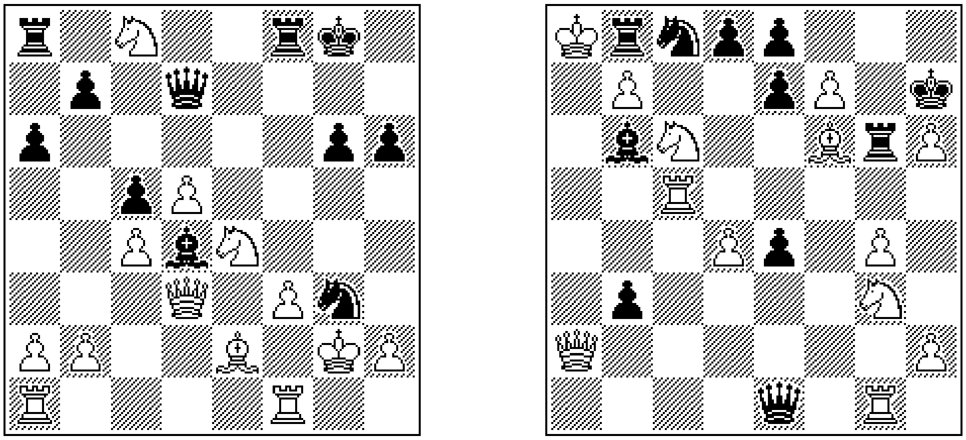


Figure 10. Rules of chunking in chess positions in conflict with Gestalt principles of perceptual organization. Pieces in group A are likely to be recalled together because of the attack relationship (knight attacks bishop), but pieces in group B are unlikely to be recalled together despite the similarity (of shape), proximity, and collinearity relationships.



Normal Position

Random Position

Figure 11. Sample stimuli from Experiment 5. (See text for explanation.)

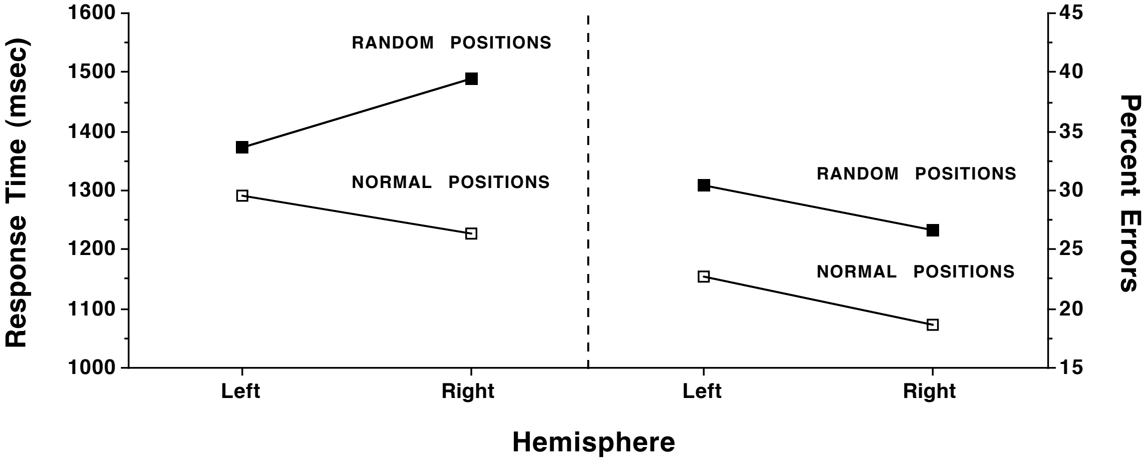
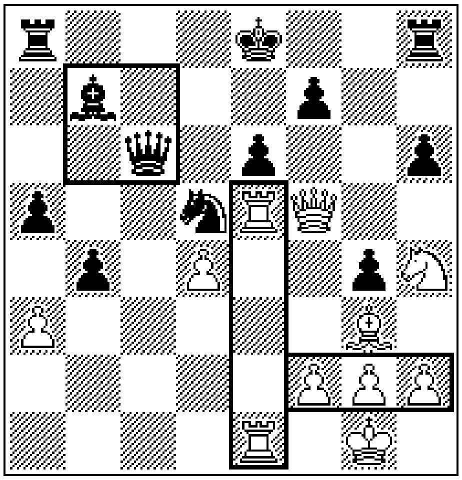
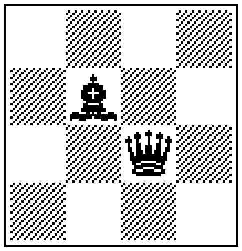


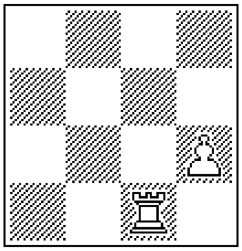
Figure 12. Results of Experiment 5: Mean response times (*left panel*) and error rates (*right panel*) for normal and random chess positions presented initially to the left and right hemispheres ("yes" responses only).



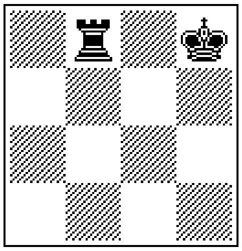
Study Position



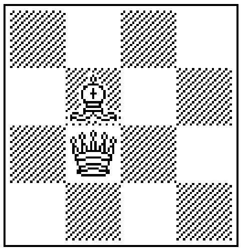
single-chunk YES



multi-chunk YES



single-chunk NO



multi-chunk NO

Figure 13. Sample stimuli from Experiment 6. (See text for explanation.)

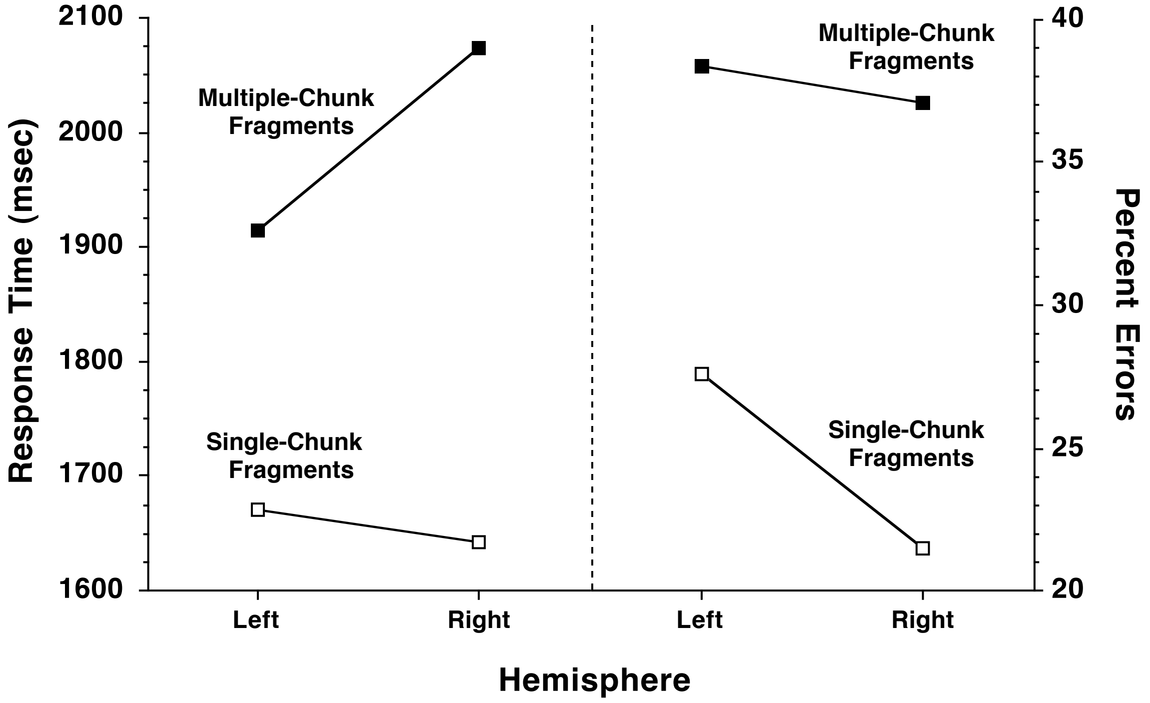


Figure 14. Results of Experiment 6: Mean response times (*left*) and error rates (*right)* for single-chunk and multiple-chunk fragments presented initially to the left and right hemispheres ("yes" responses only).

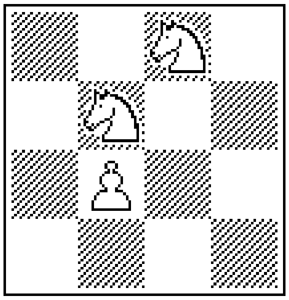


Figure 15. A fragment stimulus from Experiment 6, coded for the independent variables used in the regression analysis as follows: *Type*, multiple-chunk. *Number of pieces,* three. *Collinearity,* medium. *Proximity,* low. *Piece color similarity,* high. *Piece type similarity,* medium. *Square color similarity,* low. *Contrast,* medium.