CHAPTER 29

Eye movements and visual expertise in chess and medicine

Eyal M. Reingold and Heather Sheridan

Abstract

The chapter highlights the theoretical and applied contributions of eye movement research to the study of human expertise. Using examples drawn from the domains of chess and medicine, the chapter demonstrates that eye movements are particularly well-suited for studying two hallmarks of expert performance: the superior perceptual encoding of domain related patterns, and experts’ tacit (or implicit) domain related knowledge. Specifically, eye movement findings indicate that expertise is associated with a greater ability to process domain related visual information in terms of larger patterns of features rather than isolated features. Furthermore, in support of the role of tacit knowledge in expertise, there is evidence that the eye movements of experts may contain information that is not consciously accessible.

Introduction

To assess the usefulness of eye movement measurement for the study of expertise one need look no further than the pivotal role played by eye movement paradigms in the study of reading skill (see Rayner, 1998 for a review; Part 5 of this volume). However, there is a rapidly growing literature employing eye movement monitoring to study expertise in domains other than reading. Our survey of this literature found over 1000 relevant publications, in over a dozen skill domains. In addition to chess and medicine, which are the focus of the present chapter, these domains include art (e.g. Kozbelt and Seeley, 2007; Locher, 2006; Nodine et al., 1993), aviation (e.g. Ahlstrom and Friedman-Berg, 2006; Sarter et al., 2007), driving (e.g. Hills, 1980; Shinar, 2008; Underwood, 2007), forensics and security (e.g. Bond, 2008; Dyer et al., 2006; McCarley et al., 2004), music reading (e.g. Goolsby, 1989; Madell and Hebert, 2008; Rayner and Pollatsek, 1997), sports (e.g. Land and McLeod, 2000; Mann et al., 2007; Vickers, 1992), scientific knowledge (e.g. Jarodzka et al., 2010; Tai et al., 2006; Van Gog et al., 2005), teaching (e.g. Behets, 1996; Petrakis, 1993), and typing (e.g. Butsch, 1932; Inhoff and Gordon, 1997).

A review of these studies is beyond the scope of the present chapter. Instead we illustrate the unique contributions of eye movement studies to the study of skilled performance by reviewing findings from two domains of expertise: chess and medicine. The domain of chess was selected due to its crucial historical significance for expertise research, while the domain of medicine was chosen because of the extensive and productive use of eye movement paradigms in this domain during the
past four decades. In addition, the present chapter is focused on two fundamental aspects of skilled performance for which eye movement techniques are uniquely suited: 1) the superior perceptual encoding of domain related patterns by experts and 2) the exploration of tacit, or implicit knowledge, which constitutes a hallmark of human expertise. In the remainder of this chapter, we introduce these topics and review related studies followed by a brief concluding section.

**Expertise and superior perceptual encoding of domain related patterns**

Introductory textbooks to cognitive science typically discuss perception in the early book sections and problem solving and expertise in the final book sections to reflect the progression from ‘low-level perception’ to ‘high-level cognition’. However, one of the most fascinating and impressive aspects of skilled performance is the ability of the experienced eye to encode at a glance the essence of briefly presented stimulus material that is related to the domain of expertise (henceforward, domain related patterns). The most influential investigation of the perceptual aspects of skilled performance originated from pioneering work on expertise in chess by de Groot (1946/1965) and Chase and Simon (1973a, 1973b). Indeed, this research is often regarded as the origin of modern expertise research. Accordingly, we begin this section by briefly reviewing this work and its impact. We then review eye movement findings concerning both the superior perceptual encoding of chess related patterns by experts, and the unconscious processing of chess related patterns by experts. Later on in this chapter, the generality of these findings is assessed by reviewing related findings concerning eye movements and visual expertise in medicine.

**Perception in chess**

Chess research dates back to the beginning of modern experimental psychology during the late 1800s and early 1900s (e.g. Binet, 1894; Cleveland, 1907; Djakow et al., 1927). Simon and Chase (1973) argued that similar to the use of Drosophila (the fruit fly) as a model organism for the study of genetics, chess offers cognitive scientists an ideal task environment for the study of cognitive processes in general, and skilled performance in particular. Similarly, Newell and Simon (1972) selected chess as one of the three model tasks that they used in developing their highly influential information processing theory of human problem solving. Consistent with these suggestions, during the last century chess research has proven to be very instrumental in enhancing our understanding of human expertise and in contributing to the study of AI (for reviews see Charness, 1992; Ericsson and Charness, 1994).

Arguably, the most important contribution of chess research is in producing a major theoretical shift in the conceptualization of expertise in cognitive science away from viewing skilled performance as the product of superior general intelligence and innate talent, and toward the recognition that expertise largely reflects domain specific knowledge acquired through extensive and deliberate practice (for a review see Ericsson and Charness, 1994). This dramatic change in perspective originated from pioneering work on chess by de Groot (1946/1965) and Chase and Simon (1973a, 1973b). De Groot (1946/1965) presented chess positions briefly (2–15 s) and then removed them from view. Even after such a brief exposure the best chess players were able to reproduce the locations of the chess pieces almost perfectly (about 93% correct for positions containing about 25 pieces), and substantially better than less skilled players. In a classic study, Chase and Simon (1973a, 1973b) replicated and extended de Groot’s findings by demonstrating that after viewing chess positions for only a few seconds, chess masters were able to reproduce these positions much more accurately than less skilled players. Chase and Simon also presented chess positions with randomly rearranged chess pieces. There was little difference as a function of expertise when random board configurations were used, which indicates that the superior immediate memory performance of the skilled players was not attributable to the general superiority or
Eye movements and visual expertise in chess and medicine · 525

unique structure of their memory systems or processes. More recently, a very small but reliable advantage in recall for random configurations has been shown for expert players, though this is probably attributable to occasional presence of familiar configurations in random positions (Gobet and Simon, 1996a).

Taken together, the findings reported by de Groot and Chase and Simon suggest that chess grandmasters use efficient perceptual encoding of chess configurations to generate the most promising candidate moves and to restrict their reliance on the effortful and slow serial search through the space of possible moves. Consistent with this suggestion, both de Groot and Chase and Simon highlighted the importance of perceptual encoding of chess configurations as a key determinant of chess skill. For example, in a seminal paper entitled ‘Perception in Chess’, Chase and Simon (1973a) introduced their chunking theory of skilled performance in chess. Echoing an earlier conclusion by de Groot (1946/1965) that the efficiency of perceptual encoding processes was a more important differentiator of chess expertise than was the ability to think ahead in the search for good moves, Chase and Simon (1973a) argued ‘that the most important processes underlying chess mastery are these immediate visual-perceptual processes rather than the subsequent logical-deductive thinking processes’ (p. 215). Chase and Simon (1973a, 1973b) proposed that through extensive study and practice, expert players build up associations between perceptually recognizable chunks (i.e. groups of chess pieces related by type, colour, or role) and long-term memory structures that trigger the generation of plausible moves. Search for the best move is thereby constrained to the more promising branches in the space of possible moves from a given chess position. The size of an expert’s vocabulary of chess related configurations was initially estimated to be 50,000–100,000 chunks (Simon and Gilmarth, 1973). However, a more recent estimate puts the number of chunks at approximately 300,000 (Gobet and Simon, 2000). In addition, small perceptual chunks are most likely supplemented by larger structures termed templates (Gobet and Simon, 1996b, 1998).

Eye movements in chess: predictions and studies

An important goal of the present review is to illustrate the potential role of eye movement measurement in supplementing traditional measures of performance such as reaction time (RT), accuracy, and verbal reports as a means for investigating the perceptual aspects of skilled performance in general, and chess skill in particular. One facilitating factor for using eye movement measurement in chess is that just like words and sentences, the chess board is easily visually segmentable. In addition, if as suggested by Chase and Simon and de Groot, chess masters perceptually encode chess positions more efficiently by relying upon larger patterns of related pieces (i.e. chunks), then several predictions concerning the differences in eye movement patterns between expert and intermediate players can be made. Specifically, chess experts’ encoding of chunks rather than individual pieces should result in fewer fixations, and fixations between rather than on individual pieces. This may also imply that in any given fixation that is produced while examining structured but not random chess configurations, experts process information about a larger segment of the chessboard than less skilled players constituting an increase in the visual span as a function of expertise (the term visual span is also referred to in the literature as the perceptual span or the span of effective vision, see Jacobs, 1986; Rayner, 1998). Such a visual span advantage should also mean that experts make greater use of peripheral and parafoveal processing to extract information from a larger portion of a chessboard during an eye fixation. In addition, experts may make greater use of automatic and parallel extraction of chess relations relative to intermediate players.

Several early studies employing eye movement measurement provided weak support for the idea that perception of chess related configurations improves with skill. Tikhomirov and Poznyanskaya (1966) and Winikoff (1967) both found evidence that when chess players fixate on a chess piece, they also extract information about other pieces near the point of gaze and often move to fixate on a related piece. Based on this general process, Simon and Barenfeld (1969) devised a computer model to simulate the initial scanning patterns chess players might use when encoding a chess position. Their simulation, PERCEIVER, produced eye movement patterns that resembled those of chess players.
Reynolds (1982) and Holding (1985) re-examined the eye movement data collected by Tikhomirov and Poznyanskaya (1966), and noted that many fixations did not fall on pieces, but on empty squares. There was no report of systematic variation in the proportion of fixations on empty squares as a function of skill.

Re-analysing the work of Longman (1968), de Groot and Gobet (1996) reported no significant difference in the proportion of fixations on empty squares as a function of skill. These authors cautioned however that the negative results do not necessarily refute the chunking hypothesis. They pointed out that the crude frame-by-frame analysis of film records of eye movements and the transformation of gaze positions from a three-dimensional chessboard viewed by the players to a two-dimensional coordinate system may have resulted in the introduction of noise making it difficult to estimate the accuracy of the computed gaze position. Furthermore, de Groot and Gobet (1996) demonstrated that skilled players made more fixations along the edges of squares (28.7% of fixations) as compared with novices (13.7%), providing some indication that the skilled players may be able to encode two or more pieces in a single fixation. In addition, de Groot and Gobet (1996) concluded based on their analysis of retrospective verbal reports that the best players tended to perceive groups of pieces, rather than individual pieces.

More recently, a research programme by Reingold, Charness, and their colleagues (Charness et al., 2001; Reingold and Charness, 2005; Reingold et al., 2001a, 2001b) that employed more modern eye movement paradigms provided strong support for enhanced perceptual encoding as a function of chess expertise. Across studies Reingold, Charness, and their colleagues employed three different tasks: 1) a check detection paradigm, 2) a ‘change blindness’ flicker paradigm, and 3) a move-choice task. These tasks and the findings obtained will be discussed below.

The check detection task
Saariluoma (1985) has shown that master players can rapidly and accurately decide whether a chess piece is attacked, and do so more quickly than their less skilled counterparts. The rather simple chess relation of check detection (attack of a King) is highly salient and presents a good model for the extraction of chess-relevant relations among pieces. As shown in Fig. 29.1A, the check detection task employed by Reingold et al. (2001a) was performed using a minimized 3×3 chessboard containing a Black King and one or two potentially checking pieces. At the beginning of each trial, participants fixated the centre square of the board, a square that was always empty. A large visual span in this task may result in few if any saccades during a trial and in fixations between, rather than on individual pieces. To demonstrate that the encoding advantage of experts is related at least in part to their chess experience, rather than to a general perceptual superiority, Reingold et al. (2001a) manipulated the familiarity of the notation (symbol vs. letter) used to represent the chess pieces (see Fig. 29.2A). The symbol and letter notations were used to represent identical chess problems. However, the symbol representation is much more familiar than the letter representation. Consequently, if encoding efficiency is related to chess experience, any skill advantage should be more pronounced in the symbol than in the letter trials (i.e. a skill by notation interaction).

In order to compare the spatial distributions of gaze positions in the check detection task across the novice, intermediate and expert groups, Fig. 29.2B shows scattergrams with each dot representing an individual gaze position. An inspection of the scattergrams collapsing across all trial types (i.e. the spatial layout of chess pieces, check status, and notation), reveals a greater concentration of black pixels in the centre of the scattergram for the experts as compared to the intermediates and novices. This centre of gravity effect reflects a large disparity between skill groups in the proportion of trials without an eye movement (i.e. No-saccade trials). In such trials the gaze position remained in the centre square of the chessboard throughout the duration of the trial. For each skill group by notation type, Panel C of Fig. 29.2 displays the proportion of No-saccade trials. As can be clearly seen in this figure, only the expert group demonstrated a substantial proportion of No-saccade trials and the proportion of such trials was greater for the symbol notation than the letter notation in this group of players.
Fig. 29.1 Scan path efficiency as a function of expertise for chess players (A) and radiologists (B).

As shown in Fig. 29.2D, on trials in which eye movements occurred, experts made fewer fixations than intermediates and novices. In addition, compared to their less skilled counterparts, experts placed a smaller proportion of these fixations on pieces (Fig. 29.2E). More importantly, all skill related differences were more pronounced in the symbol than the letter notation. Thus, consistent with Chase and Simon’s chunking hypothesis, in the check detection task chess experts made fewer fixations and placed a greater proportion of fixations between individual pieces, rather than on pieces. The magnitude of these effects was stronger for the more familiar symbol notation than for the letter notation, demonstrating that the experts’ encoding advantage is related at least in part to their chess experience, rather than to a general perceptual superiority.
Fig. 29.2 The check detection task from Reingold et al. (2001a). The notation manipulation (A). Scattergrams of gaze positions in the check detection task (B) the capital letter ‘A’ represents the position of an attacker piece and ‘K’ represents the position of the King. Proportion of No-saccade trials (C), number of fixations (D), and proportion of fixations on pieces (Panel E). Eyal M. Reingold, Neil Charness, Marc Pomplun, and Dave M. Stampe, *Psychological Science*, 12(1), copyright © 2001 by Sage Publications. Reprinted by Permission of SAGE Publications.
The change blindness flicker paradigm

The flicker paradigm was introduced by Rensink et al. (1997). Reingold et al. (2001a) used two types of configurations: chess configurations (with 20 chess pieces in each) selected from a large database of chess games, and random configurations, which were created by repeatedly and randomly exchanging pieces in the chess configurations. Thus, random positions maintained the same spatial configuration but destroyed the chess relation information. Each random or chess configuration was modified by changing the identity but not the colour of a single piece to create a modified display (see Fig. 29.3A). In each trial, images of the original and modified board configurations were displayed sequentially and alternated repeatedly with a blank interval between each pair of configurations. Each variant of the configuration (i.e., original, modified) was presented for 1000 ms, with the display blanking for 100 ms between each alternation. As soon as participants detected the changing piece (the target), they ended the trial by pressing a button and naming the alternating pieces. Previous research indicated that participants are surprisingly poor at change detection in the flicker paradigm, a phenomenon termed 'change blindness' (Rensink et al., 1997; for a review see Simons and

![Fig. 29.3 Illustration of the gaze-contingent flicker paradigm from Reingold et al. (2001a). A) Shows an original (bottom left) and a modified (top right) chess configuration (with a box indicating the changed piece) and with chess pieces outside the window being replaced by blobs masking their identity and colour. B) Area of visual span (number of squares). Eyal M. Reingold, Neil Charness, Marc Pomplun, and Dave M. Stampe, *Psychological Science, 12*(1), copyright © 2001 by Sage Publications. Reprinted by Permission of SAGE Publications.](image)
Levin, 1997). It was predicted that when processing chess configurations, but not random configurations, chess experts would demonstrate larger visual spans than less skilled players.

In this task the visual span as a function of chess skill (novice vs. intermediate vs. expert) and configuration type (chess configuration vs. random configuration) was measured using a gaze-contingent window technique (e.g. McConkie and Rayner, 1975; see Rayner, 1998, for a review). As shown in Fig. 29.3A, a gaze-contingent window requires obscuring the identity of all chess pieces except those within a certain 'window' that is continually centred on the participant's gaze position. The pieces outside a circular, gaze-centred window were replaced with grey blobs masking the actual colours and shapes. The participant's visual span was measured by varying the size of the window over successive trials and determining the smallest possible window that did not significantly differ from the participant's normative RT criteria. These criteria were established separately for chess configuration and random configuration by using baseline trials in which the entire display was visible (i.e. No-window trials). Note that this change detection task required no chess knowledge and consequently Reingold et al. (2001a) were able to explore visual span across a broad range of chess skill stretching from novice to master. For each skill group by configuration type, Panel B of Fig. 29.3 displays the visual span results. Experts' span area for chess configurations was dramatically larger than all other skill group by configuration type cells, which in turn did not differ from each other. Thus, consistent with Chase and Simon's hypothesis, the increase in visual span area which characterizes expert performance on trials with chess, but not random configurations, clearly indicates an encoding advantage attributable to chess experience, rather than to a general perceptual or memory superiority.

The move-choice task

Given the strong support for enhanced perceptual encoding as a function of chess expertise obtained in the tasks reviewed above, Charness et al. (2001) and Reingold and Charness (2005), attempted to extend these findings to the more ecologically valid task of choosing the best move with full chessboard displays (henceforth, the move-choice task). Focusing on the perceptual encoding phase during move-choice trials, Charness et al. (2001) restricted their analysis to the first five fixations in each trial (approximately the first 1–2 s). Consistent with the check detection findings, experts produced a greater proportion of fixations on empty squares than intermediates (experts: \( M = 0.52 \); intermediates: \( M = 0.41 \)). In addition, consistent with de Groot and Gobet (1996), among fixations on pieces, experts produced a greater proportion of fixations on salient pieces (i.e. active pieces that were relevant to generating the best move) than intermediates (experts: \( M = 0.80 \); intermediates: \( M = 0.64 \)). Piece saliency was determined by asking two international masters to classify pieces in each position used in the experiment as salient or non-salient. Thus, as indicated by the spatial distribution of early fixations, experts processed larger patterns or chunks and such processing of global position information might have enabled them to be remarkably efficient in rapidly identifying task relevant pieces and configurations.

Whereas in Charness et al. (2001) the focus was on the first five eye fixations during the performance of the move-choice task, in the follow-up experiment Reingold and Charness (2005) recorded fixations during the first 10 s in each trial. They hypothesized that an examination of changes in the number and duration of fixations, which might occur as the trial progresses, would be potentially useful in distinguishing between perceptual encoding and problem solving or solution retrieval and evaluation. Specifically, perceptual encoding was expected to involve shorter fixations and consequently a greater number of fixations in a given time interval than problem solving. Reingold and Charness (2005) were also interested in the proportion of fixations with durations greater than 500 ms Such fixations have been identified previously as reflecting visual problem solving and evaluation (e.g. Nodine et al., 1978).

Fig. 29.4A shows scattergrams aggregating all fixations in the first 10 s for both intermediates (left panel) and experts (right panel) for one of the positions used in the experiment. Each circle represents an individual fixation, and the diameter of the circle increases as a function of an increase in
Fig. 29.4 Scattergrams aggregating all fixations in the first 10 s across intermediates (left) and experts (right) for one of the positions used in the move choice task in Reingold and Charness (2005) (A). Each circle represents an individual fixation, and the diameter of the circle increases as a function of an increase in fixation duration. Best move for this position = White Queen takes Pawn at h5 check, Black Rook takes White Queen, White Bishop moves to g6 mate. Number of fixations (left), and the proportion of fixations longer than 500 ms (right) in the move choice task by skill and interval (first 5-s interval vs. second 5-s interval) (B). Reproduced from G. Underwood, Cognitive Processes in Eye Guidance, 2005, with permission from Oxford University Press.

fixation duration. As can be clearly seen by comparing the scattergrams, consistent with the chunking hypothesis and the findings of Charness et al. (2001), experts produced a greater proportion of fixations on empty squares than intermediates (experts: M = 0.55; intermediates: M = 0.43). In addition, as indicated by a comparison of the relative size of the circles across scattergrams, experts clearly produced a higher proportion of longer fixations than intermediates. Reingold and Charness (2005) divided the 10-s period of eye movement recording in the beginning of each trial into two 5-s intervals. For each interval, the mean number of fixations and the proportion of fixations with durations greater than 500 ms were computed. Fig. 29.4B displays these two dependent variables by skill group and interval. As can be seen in this figure, the pattern of performance is qualitatively different across experts and intermediates. Specifically, for intermediates there was no difference across intervals in
the number of fixations and in the proportion of long fixations (i.e. >500 ms). In marked contrast, experts produced substantially fewer fixations and a much greater proportion of long fixations as the trial progressed. This indicates that during the second 5-s interval in a trial experts started engaging in problem solving, whereas intermediates were still perceptually encoding the chess configurations. This provides further support for the hypothesis of enhanced efficiency of pattern recognition processes in encoding chess configurations as a function of chess expertise.

Unconscious processing of chess related patterns by experts

The results reviewed above suggest that it is likely that the main perceptual advantage for experts is not in the identification of single chess pieces and board locations, but rather in the extraction of relational information between pieces. This was powerfully demonstrated by the strong skill effects on the area of the visual span obtained with actual chess configurations (i.e. where relational information is intact), coupled with the absence of skill effects on span size obtained with random configurations (i.e. where relational information is broken down). Based on this dramatic demonstration of larger visual span in chess experts, in subsequent studies Reingold et al. (2001b) and Reingold and Charness (2005) proposed that one possible mechanism that might allow chess masters to process chess configurations more efficiently is automatic and parallel extraction of several chess relations that together constitute a meaningful chunk. These studies contrasted the standard check detection trials with two attackers with trials in which one of two attackers was cued (coloured). In the latter condition, the task was to determine if the cued attacker was checking the King while ignoring the other attacker. Three different trials in which the correct response was No were contrasted: 1) a standard No-Cue condition (i.e. no cueing), with 2 non-checking attackers, 2) a cued non-checking attacker that appeared together with another non-checking attacker (i.e. a congruent condition), and 3) a cued non-checking attacker that appeared together with a checking attacker (i.e. incongruent condition). Note that all of these trials are No trials even though the Incongruent condition contains a checking attacker. That is, in the Incongruent condition the semantics of the cued chess relation (i.e. no check) is inconsistent with the semantics of the configuration as a whole (i.e. check). Serial processing of chess relations will manifest as faster processing (e.g. faster RTs, fewer fixations) in the Congruent condition, than in the No-Cue condition, as the cueing constrains the search space. In contrast, parallel processing of chess relations should result in no benefit from cueing in the Congruent condition. In addition, if parallel processing of chess relations occurs, cueing should produce slower processing in the Incongruent, than in the Congruent condition, demonstrating Stroop-like interference. While novices and intermediates displayed serial processing of chess relations, experts showed a parallel processing pattern including a Stroop like interference effect. Such an interference effect constitutes the ‘gold standard’ for demonstrating automaticity and an absence of conscious control. In other words, the experts, but not their less skilled counterparts were unable to avoid processing the global pattern even when such processing was detrimental to task performance.

Eye movements have also revealed evidence of unconscious processing in a chess study concerning the Einstellung (set) effect (Bilalić et al., 2008). This effect refers to the finding that in many problem-solving situations, initially generating a less optimal solution can prevent a better solution from coming to mind. In the case of chess, Bilalić et al. (2008) observed that when expert players (Candidate Masters) are asked to find the fastest way to win a game, they might miss the best solution (checkmate in three moves) when a familiar, less optimal solution is also present (checkmate in five moves). To study the mechanisms underlying this effect, Bilalić et al. (2008) compared the percentage of time that players spent looking at the squares that were important for the optimal versus the less optimal solution. In retrospective reports, the players claimed that they had continued to search for a faster solution after discovering the five-move solution. However, their eye movement record showed that throughout the problem-solving period they spent a greater proportion of time looking at the squares associated with the less optimal than the optimal solution. This finding suggests that the Einstellung (set) effect may work by directing the problem-solvers’ attention towards evidence which is consistent with their initial solution and preventing them from considering evidence that is consistent
with new solutions. Furthermore, the inconsistencies between the retrospective reports given by the players and their pattern of eye movements suggest that their eye movements reflected information that was not accessible to awareness.

Superior perceptual encoding and visual expertise in medicine

In medical domains such as radiology, pathology, dermatology, and minimally invasive surgery, there is a growing reliance on imaging equipment and the development of specialized visual expertise. Like experts in other domains, medical experts exhibit superior performance as indicated by their faster decision times and greater accuracy on domain related tasks (for a review, see Nodine and Mello-Thoms, 2000). Importantly for the present context, the application of eye tracking to the study of visual expertise in medicine (primarily in the field of radiology) has revealed a number of differences between experts and novices. As indicated in the review below, although eye movement studies in chess and medicine constitute independent research efforts with almost no overlap or cross citation, there is substantial theoretical and empirical convergence across these two domains. Table 29.1 contains brief summaries (in chronological order) of studies that directly compared the eye movements of experts and novices in order to study medical expertise. Taken together, many of these studies suggest visual expertise in medicine involves superior encoding of domain related configurations (often referred to as a global processing advantage) by experts (see Fig. 29.1B for an illustration of increased efficiency of scan paths as a function of skill).

To explain the superior encoding of domain related patterns by medical experts, models, such as the global-focal search model (Nodine and Kundel, 1987) and the two-stage detection model (Swensson, 1980), have incorporated a global processing component. Both of these models postulate that experts have more efficient scan paths because they can simultaneously process visual information from across a wide field of vision (i.e. a larger visual span). According to the global-focal search model, upon viewing an image, experts quickly obtain a global impression, which consists of a comparison between the contents of the image and the expert’s schema. The term schema refers to the expert’s knowledge about the overall visual patterns that are associated with abnormal and normal radiographs. During the global impression, the expert makes note of both perturbations, which are deviations from the expert’s schema, and of potential abnormalities. These perturbations and the potential abnormalities are then scanned with the fovea. Thus, the expert radiologist’s knowledge about the visual patterns typically associated with normal and abnormal anatomic structures allows them to quickly pinpoint suspicious regions. As pointed out by Guderman et al. (2001), this explanation of how prior experience shapes search patterns in radiology is similar to the concept of ‘chunking’ (Miller, 1956). As such the global-focal search model is also reminiscent of Chase and Simon’s chunking theory of skilled performance in chess. Both models argue that given that experts have a vocabulary of domain specific visual patterns built up from their prior experiences in the domain, they are able to evaluate larger constellations of features, instead of only perceiving individual features.

Similarly, the two-stage detection model (Swensson, 1980) incorporates the idea that radiologists can quickly evaluate large regions of an image. This model assumes that expert radiologists have developed perceptual mechanisms, which act as an initial filter that automatically identifies features that require further examination. With experience, these visual mechanisms have been trained to filter out normal structures, in order to direct the radiologist’s attention to structures that are likely to be abnormalities. Thus, the notion that experts engage in global processing in order to flag potential abnormalities is captured by both the global-focal model (Nodine and Kundel, 1987) and by the two-stage detection model (Swensson, 1980). As outlined below, a variety of evidence is consistent with these two models.

Early evidence that radiologists engage in global processing was provided by the brief exposure studies, which are sometimes referred to as ‘flash studies’. When an image is presented for a brief period of time, radiologists can still identify a large number of abnormalities (Carmody et al., 1980a; 1981; Gale et al., 1990; Kundel and Nodine, 1975; Mugglestone et al., 1995; Oestmann et al., 1988).
<table>
<thead>
<tr>
<th>Reference</th>
<th>Range of expertise</th>
<th>Task</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kundel and La Follette (1972)</td>
<td>Laypersons, medical students, residents, radiologists</td>
<td>Lung nodules in chest radiographs</td>
<td>The radiologists showed more efficient scan paths and required fewer fixations to reach the abnormalities</td>
</tr>
<tr>
<td>Kundel (1974, 1983)</td>
<td>Laypersons, radiologists</td>
<td>Lung nodules in chest radiographs</td>
<td>The radiologists spent more time fixating the abnormalities in abnormal films, and more time fixating areas that were more likely contain abnormalities in the normal films</td>
</tr>
<tr>
<td>Carmody et al. (1984)</td>
<td>Radiology residents, radiology instructors</td>
<td>Lung nodules in chest radiographs</td>
<td>This study reports a discrepancy between verbal reports and eye movements: radiology experts reported that they frequently used bilateral comparison scans (alternating between the same sections on the two lungs). In contrast, radiologists’ eye movements showed that this type of scanning occurred only 4% of the time</td>
</tr>
<tr>
<td>Krupinski (1996a, 1996b, 2000)</td>
<td>Radiology residents, Radiologists</td>
<td>Mammograms</td>
<td>Radiologists had shorter overall viewing times, faster times to first fixation on abnormality, more efficient scan patterns, covered less of the image, and had longer false-negative durations</td>
</tr>
<tr>
<td>Nodine et al. (1996a, 1996b)</td>
<td>Laypersons, residents, technologists, mammographers</td>
<td>Mammograms</td>
<td>Mammographers had faster times to first fixation on the abnormality, and longer fixation times on the lesion</td>
</tr>
<tr>
<td>Nodine and Kundel (1997); Nodine and Krupinski (1998)</td>
<td>Laymen, radiologists</td>
<td>Lung nodules in chest radiographs</td>
<td>Radiologists had faster times to first fixation on the abnormalities in the chest images, but not in control visual search tasks</td>
</tr>
<tr>
<td>Wooding et al. (1999)</td>
<td>Mean experience in months: 0 (laymen), 2.3 (novices), 16.5 (trainees), 90 (radiologists)</td>
<td>Lung nodules in chest radiographs</td>
<td>The fixation patterns of trainees show the least within-group consistency and the least amount of similarity to radiologists, suggesting that trainees go through a developmental phase characterized by more idiosyncratic patterns of attention allocation and eye movements</td>
</tr>
<tr>
<td>Krupinski (2000)</td>
<td>Residents, mammographers</td>
<td>Mammograms</td>
<td>Observers’ eye movements were monitored before and after computer-aided detection (CAD) prompts, which identified locations containing suspicious features that were likely to be a lesion. Prior to the prompt, the mammographers had longer false-negative durations than the residents. After the prompt, the residents had longer durations for all four response outcomes</td>
</tr>
<tr>
<td>Nodine et al. (2002)</td>
<td>Trainees, mammographers</td>
<td>Mammograms</td>
<td>Mammographers were more likely to fixate on the lesion for 1000ms or greater. Both groups showed longer fixations for false-negative than for true-negative responses and this difference was numerically larger for the mammographers</td>
</tr>
<tr>
<td>Reference</td>
<td>Range of expertise</td>
<td>Task</td>
<td>Key findings</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------</td>
<td>---------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mello-Thoms et al. (2002); Mello-Thoms (2003)</td>
<td>Residents, mammographers</td>
<td>Mammograms</td>
<td>Mammographers were more consistent than residents in terms of their visual search and decision-making strategies. For residents, but not mammographers, there were differences in local spatial frequency information between missed lesions that attracted visual attention, and those that did not, indicating that the residents’ search strategy was more influenced by the local saliency of the lesions whereas the experts were better able to make use of global patterns.</td>
</tr>
<tr>
<td>Manning et al. (2003, 2006a)</td>
<td>Novices, radiographers, radiologists</td>
<td>Lung nodules in chest radiographs</td>
<td>Increased experience resulted in longer saccades amplitudes, fewer fixations, less coverage of the image, and faster decision times.</td>
</tr>
<tr>
<td>Law et al. (2004)</td>
<td>Novices, surgeons</td>
<td>Laparoscopic surgery simulation</td>
<td>While performing an aiming task that involved touching a small target with the tip of a tool, the experts spent more time fixating on the target, less time fixating on the tool, and had fewer saccades between the tool and the target.</td>
</tr>
<tr>
<td>Kundel and Nodine (2004)</td>
<td>Residents, mammography fellows, mammographers</td>
<td>Mammograms</td>
<td>Mammographers had longer false-negative durations, were more likely to fixate on lesions, and had faster times to first fixation on abnormalities.</td>
</tr>
<tr>
<td>Manning et al. (2004, 2006b)</td>
<td>Novices, radiographers, radiologists</td>
<td>Lung nodules in chest radiographs</td>
<td>All the groups showed longer fixation times for false-negative than true-negative responses, and this effect tended to be larger for the inexperienced groups. The experienced groups covered less of the image.</td>
</tr>
<tr>
<td>Donovan et al. (2005)</td>
<td>Laymen, first-year students, third-year students, experts</td>
<td>Wrist fractures</td>
<td>Perceptual feedback about eye movements resulted in decreased performance by experienced observers and showed a trend towards improving the performance of less experienced observers.</td>
</tr>
<tr>
<td>Kocak et al. (2005)</td>
<td>Novice, intermediate, expert</td>
<td>Laparoscopic surgery simulation</td>
<td>As expertise increased, there was a decrease in the saccadic rate and an associated increase in mean fixation times, and there was a trend towards larger saccadic amplitudes.</td>
</tr>
<tr>
<td>Krupinski (2005)</td>
<td>Residents, mammographers,</td>
<td>Mammograms</td>
<td>The mammographers had faster times to first fixation on abnormality, and this effect was greater in magnitude for the more subtle lesions. For both groups, lesions associated with true-positive outcomes were fixated for longer than lesions associated with false-negative outcomes, and subtle lesions were fixated for longer than more obvious lesions.</td>
</tr>
</tbody>
</table>
### Table 29.1 (continued) Summary of studies of eye movements and medical expertise

<table>
<thead>
<tr>
<th>Reference</th>
<th>Range of expertise</th>
<th>Task</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Krupinski et al.</td>
<td>Medical students, residents,</td>
<td>Telepathology virtual slides</td>
<td>The observers' task was to view a low magnification image, and to decide on the top three locations that they would zoom into if they were to continue to examine the slide. The pathologists showed a more efficient scan path with shorter overall viewing times, fewer fixations, fewer saccades and longer saccades. They did not fixate at all, significantly more frequently than the trainees, on some of their preferred zoom locations, indicating greater use of parafoveal and peripheral vision to extract global information.</td>
</tr>
<tr>
<td>(2006)</td>
<td>pathologists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burgert et al.</td>
<td>13 surgeons with varying levels</td>
<td>Neck dissection planning</td>
<td>Experienced surgeons showed fewer saccades and longer fixations on areas of interest, indicating more efficient integration of 3-dimensional global information.</td>
</tr>
<tr>
<td>(2007)</td>
<td>of expertise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kundel et al.</td>
<td>Observers varied in experience,</td>
<td>Mammograms</td>
<td>The more accurate performers had faster times to first fixation on the abnormality, and were more likely to exhibit a long saccade towards the abnormality immediately upon viewing the image, indicating greater peripheral and parafoveal processing of global information.</td>
</tr>
<tr>
<td>(2007)</td>
<td>and expertise was defined based</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>on detection accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leong et al.</td>
<td>Expertise varied widely from</td>
<td>Fracture detection (shoulder,</td>
<td>For true-positive trials, the more experienced observers spent a smaller proportion of time fixating the fracture site but there was no difference between the groups on false-negative trials. The more experienced observers showed greater consistency in their search patterns. The experts, but not the novices, displayed distinct stages in eye movement search patterns.</td>
</tr>
<tr>
<td>(2007)</td>
<td>officers to surgeons and</td>
<td>hand, and knee)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>radiologists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donovan et al.</td>
<td>Laymen, level 1 students, level</td>
<td>Lung nodules in chest</td>
<td>Perceptual feedback about eye movements produced improvement across all levels of expertise, with the greatest improvement shown by the level 1 group.</td>
</tr>
<tr>
<td>(2008)</td>
<td>2 students, experts</td>
<td>radiographs</td>
<td></td>
</tr>
<tr>
<td>Litchfield et al.</td>
<td>Novice radiographers, expert</td>
<td>Lung nodules in chest</td>
<td>Both novice and expert radiographers show improved performance after viewing the eye tracking record of another observer who had examined the same image, but the improvements were more pronounced for the novices than the experts.</td>
</tr>
<tr>
<td>(2008)</td>
<td>radiographers</td>
<td>radiographs</td>
<td></td>
</tr>
</tbody>
</table>

For example, after viewing chest films for 200 ms, expert radiologists were able to detect 70% of the abnormalities, compared with 97% true-positives with unlimited viewing conditions (Kundel and Nodine, 1975). Similarly, expert mammographers were above chance at identifying abnormalities under flash viewing conditions, although they tended to miss the more subtle abnormalities (Mugglestone et al., 1995). To determine the limits of performance under brief exposure conditions, Carmody et al. (1980a), systematically varied the distance between the observers’ point of fixation and the location of a chest nodule. Although accuracy decreased as the distance increased, it was found that a radiologist could detect some nodules that were 15° away, and a less experienced film
reader could detect some nodules that were 10° away. Given that brief exposure conditions preclude eye movements toward the abnormalities, it appears that radiologists can identify abnormalities without foveal vision.

The evidence from brief exposure studies was later supplemented by a variety of eye movement findings. One prominent finding is that expert radiologists exhibit more efficient scan paths than novices (e.g. Krupinski 1996a, 1996b, 2000; Krupinski et al., 2006; Kundel and La Follette 1972), suggesting that they can use information that is outside of the fovea to guide their search. Figure 29.1B (taken from Kundel and La Follette, 1972; see also Kundel, 2007) shows examples of the range of scan paths exhibited by medical students, residents, and expert radiologists. As shown in this figure, expert radiologists often exhibit a circumferential scan pattern (Kundel and Wright, 1969) that involves making a wide sweep around the image with long saccades and a small number of widely spaced fixations. An efficient scan path entails quickly discovering abnormalities while at the same time covering less area with the fovea. Relative to novices, experts exhibit a number of the characteristics associated with efficient scan paths, including fewer fixations (Krupinski et al., 2006; Manning et al., 2003, 2006a), fewer saccades (Burgert et al., 2007; Kocak et al., 2003; Krupinski et al., 2006; Law et al., 2004), longer saccades (Kocak et al., 2005; Krupinski et al., 2006; Manning et al., 2003; Manning et al., 2006a), less coverage of the image (Krupinski; 1996a, 1996b, 2000; Manning, et al., 2003, 2004, 2006a) arriving at the abnormality faster (Krupinski; 1996a, 1996b, 2000, 2005; Kundel and Nodine, 2004; Kundel et al., 2007; Nodine et al., 1996a, 1996b), and spending a greater proportion of time fixating on abnormalities (Kundel, 1974; Kundel and La Follette, 1972, Kundel and Nodine, 1983, 2004; Nodine et al., 1996a, 1996b, 2002; but see Leong et al., 2007).

In further support of the models, when the number of abnormalities reported is plotted against time, experts display a rapid reporting phase, followed by a more gradual reporting phase (Christensen et al., 1981; Nodine et al., 2002). This is because the expert can quickly identify many abnormalities with peripheral and parafoveal vision, before engaging in a slower search for the more subtle abnormalities that require foveal vision. In contrast, less experienced observers show a more constant rate of reporting instead of the rapid and slow reporting phases, because the global mode of processing is less developed for them. Given that novices are unable to identify as many abnormalities with their peripheral and parafoveal vision, they must instead examine a large proportion of the image with their fovea.

The rapid reporting phase shown by experts is consistent with eye movement findings that experts fixate many abnormalities within 1 second of viewing the image (e.g. Kundel et al., 2008). As 1 s is not enough time for the entire image to be scanned with foveal vision, the short times to the first fixation on the abnormality provide strong evidence experts use parafoveal and peripheral vision to locate abnormalities. Furthermore, relative to novices, experts have faster times to their first fixation on an abnormality (Krupinski 1996a, 1996b, 2000, 2005; Kundel and Nodine, 2004; Kundel et al., 2007; Nodine et al., 1996a, 1996b). For example, Kundel et al. (2007) showed that the time to fixate the abnormality decreased as the observer’s accuracy increased, and the more accurate observers were more likely to exhibit a long saccade towards the abnormality immediately upon viewing the mammogram.

Gaze-contingent window experiments have strongly supported the role of global processing. As in chess research, the gaze-contingent window paradigm has been used to selectively vary the extent to which information is accessible by foveal, parafoveal and peripheral vision. Chest nodules were fixated faster when they were visible by parafoveal and peripheral vision, compared to a condition in which the nodule was only visible when fixated directly (within 3.5 or 5.25°) (Kundel et al., 1991). Similarly, when nodules are only visible within a central gaze-contingent window, the time it takes to first fixate the nodule decreases as the window size increases (Kundel et al., 1984). Thus, removing parafoveal and peripheral information decreases search efficiency, as indexed by the amount of time required to fixate the abnormality. Building on these findings, future research could employ the gaze-contingent window paradigm to directly compare the size of the perceptual span for medical experts and novices for both domain-relevant and non-relevant visual information.

The gaze-contingent window findings seem to be consistent with an earlier study that compared a full-image viewing condition to a segmented viewing condition, in which a chest image was divided
into six segments that were viewed one at a time (Carmody et al., 1980b). The segmented condition resulted in lower accuracy, suggesting that experts benefited from global viewing conditions. Similarly, Carmody (1984) reported decreased accuracy in a constrained viewing condition that prevented participants from making visual comparisons between different regions of the image. In line with these findings, instructing radiologists to focus on particular regions or features diminishes accuracy in comparison to a free search condition (Swensson et al., 1982; 1985). Overall, these studies provide evidence that radiology experts perform best under free search conditions that allow them to use a wide field of view to guide their search. As discussed previously, a wide field of view may improve efficiency by directing observers to abnormalities in their peripheral and parafoveal vision. However, global processing may also improve accuracy by facilitating comparisons between different regions of the same image. It has been previously argued that comparison scans (alternating between different regions with foveal vision) play an important role in visual search (Carmody, 1984; Carmody et al., 1980b), but it is possible that some visual comparisons rely on global processing. In free search conditions, radiologists may use their parafoveal and peripheral vision to detect asymmetries between different parts of the radiograph, and segmented viewing conditions could interfere with this type of processing.

Similar to the above findings from the field of radiology, there is evidence that medical experts in other fields are able to use parafoveal and peripheral information to improve the efficiency of their performance. For example, a recent study examined the eye movements of medical students, pathology residents, and practicing pathologists while they viewed telepathology virtual slides (Krupinski et al., 2006). The observers’ task was to select three locations that they would zoom into further if they were to continue to examine the slide. While selecting the three locations, the pathologists showed a more efficient scan path with fewer fixations and longer saccades. Furthermore, the experts did not fixate at all, significantly more frequently than the trainees, on some of their preferred zoom locations, suggesting that they used parafoveal and peripheral vision to extract global information.

These observations of pathology experts are consistent with reports that surgeons show fewer saccades than novices during laparoscopic surgery simulation tasks (Kocak et al., 2005; Law et al., 2004). For example, while performing a task that involved touching a small target with the tip of a tool, the experts spent more time fixating on the target, less time fixating on the tool, and had fewer saccades between the tool and the target (Law et al., 2004). This indicates that the expert surgeons were able to direct the tool using peripheral and parafoveal vision alone. Thus, the global processing advantage shown by medical experts seems to span a number of different fields, including radiology, pathology, and minimally invasive surgery.

Three main parallels can be drawn between the global processing advantage shown by medical experts, and the larger visual span shown by chess experts. First, in both cases the perceptual encoding advantage appears to be domain specific. Radiologists do not perform better than novices when tested with control visual search tasks that involved searching for the character WALDO and searching for the word NINA (Nodine and Krupinski, 1998; Nodine and Kundel, 1997), and a comparative visual search task that more closely mimics radiology tasks showed a similar pattern of results between radiologists and laymen (Moise et al., 2005). Second, chess expertise is analogous to expertise in medical diagnosis because both forms of expertise involve extensive, domain specific knowledge of visual configurations (Wood, 1999). This knowledge allows experts to ‘chunk’ together domain specific information such that they can recognize patterns instead of only seeing individual features (Gunderman et al., 2001). Furthermore, it is likely that it is necessary to build up this vocabulary of domain related visual knowledge in order to facilitate the global mode of processing. Third, for both chess players and medical experts, it is possible that not all of this knowledge is accessible to conscious awareness (Heiberg Engel, 2008; Norman et al., 1992). The remainder of this chapter will focus on eye movement findings that suggest that there is an unconscious component to expertise.

Expertise and tacit or implicit knowledge

One of the prominent modern philosophers to extensively theorise about implicit or tacit knowledge was Michael Polanyi. In describing the gradual acquisition of the skill with which a blind person uses
a stick to navigate, Polanyi provides a very eloquent portrayal of non-analytic implicit learning generating implicit knowledge:

Someone using a stick for the first time to feel his way in the dark, will at first feel its impact against his palm and fingers when the stick hits an object. But as he learns to use the stick effectively, a transformation of these jerks will take place into a feeling of the point of the stick touching an object; the user of the stick is no longer attending then to the meaningless jerks in his hand but attends from them to their meaning at the far end of the stick.

(Italics in original; Polanyi, 1969, p.145.)

With practice there is a gradual shift from attending to the manipulation and placement of the stick, to focusing on the objects in the environment which are being probed by it. Just as we are only marginally aware of the movements of our limbs during a leisurely walk (our attention is focused on the sights and sounds around us), a blind person skilled at navigating with a stick might have only a subsidiary awareness of its placement and manipulation. In a sense, the stick is being assimilated as an extension of the skilled user's body.

As the above examples illustrate, implicit knowledge may be considered an integral part of expertise. This can be powerfully demonstrated by considering Expert systems, a subarea of artificial intelligence. Practitioners in this area attempt to create computer programs that mimic the performance of human experts. A prerequisite of such an enterprise is an explicit representation of the rules and strategies used by experts to achieve skilled performance (i.e. domain specific knowledge). However, extracting the knowledge of experts turned out to be much more difficult than first thought. Basing the programs on rules experts claimed they were using resulted in inferior performance. It is not that experts were uncooperative; rather, they knew much more than they could verbalize. For the purpose of creating expert systems it is important to explicitly represent not only experts' declarative knowledge (i.e. knowing that), but also their procedural knowledge (i.e. knowing how). Unfortunately, such procedural knowledge is largely implicit (i.e. not easily verbalized or consciously accessed). To overcome this problem, designers of expert systems sometimes referred to as knowledge engineers interrogate experts as they undertake a variety of carefully selected skill related examples. Critics of expert systems point out that this effort is only partially successful at making the implicit knowledge of experts explicit. Accordingly, they argue that the performance of expert systems cannot rival that of the best human experts, and that there are restrictions on the nature and scope of the domains for which expert systems can be developed.

Eye movements have the potential to contribute to the study of expertise by revealing information about ongoing cognitive processing that is not consciously accessible to the expert. Previously, in the chess section of this chapter, we briefly discussed evidence for automatic and parallel extraction of chess related patterns in the absence of conscious control (Reingold and Charness, 2005; Reingold et al., 2001b) and evidence that chess players' eye movements provide information that is not contained in their retrospective reports (Bilali et al., 2008). We now focus on the case of medical expertise, in order to highlight the consistent finding that even when radiologists fail to report abnormalities, their eye movements still differentiate between the missed abnormalities and abnormality-free areas (for a review, see Krupinski et al., 1998; Krupinski and Borah, 2006). We will argue that both chess and medicine findings suggest that eye tracking is uniquely suited for studying the unconscious component of expertise.

Eye movements, implicit knowledge and radiology

In the field of radiology, it has been reported that inter-rater variability is high and abnormalities are missed as frequently as 30% of the time (e.g. Austin et al., 1992; Bird et al., 1992; Birkelo, Chamberlain et al., 1947; Guiss and Kuenstler, 1960). Eye tracking has been used to investigate why many abnormalities are missed even though they are visible in retrospect. Misses (referred to as false-negative responses) are frequently classified as either scanning errors, recognition errors, or decision-making errors (Kundel et al., 1978). Scanning errors occur when the radiologist does not fixate near
the abnormality. Recognition errors occur when the region containing the abnormality is fixated for a short period of time, indicating a failure to recognize the presence of a potential abnormality. Decision-making errors occur when the abnormality was fixated for a long period of time, indicating that the observer might have recognized the presence of a potential abnormality, but they incorrectly decided that the region was normal. Typically, false-negative responses are categorized as recognition errors if the cumulative cluster duration on the abnormality is less than a threshold value and as decision-making errors if the cumulative cluster duration is above the threshold value. For example, Kundel et al. (1978) used an 800-ms threshold value and found that 30% of false-negative responses were considered scanning errors, 25% were recognition errors, and 45% were decision-making errors. Subsequent research has confirmed that a substantial proportion of false-negative responses can be categorized on this basis as a failure of decision-making (e.g. Berbaum et al., 1996, 2001; Krupinski and Nishikawa, 1997; Kundel et al., 1989; Manning et al., 2004; Manning et al., 2006b). Furthermore, an additional source of false-negative errors that has been investigated extensively is the satisfaction of search (SOS) effect (Berbaum et al., 2000; Berbaum et al., 1990, 2010), which is the finding that radiologists are less accurate at detecting subtle abnormalities if additional abnormalities are present on the same radiograph.

However, regardless of how false-negative errors are classified, it is clear that missed abnormalities are fixated for a prolonged period of time relative to abnormality-free regions. There are two reasons why this eye movement finding might constitute an example of implicit knowledge. First, radiologists adopt a lenient response criterion (Scheff, 1963) for detecting abnormalities because the costs associated with missed abnormalities (the critical importance of early detection in improving outcomes) far exceed the costs associated with false alarms (patient anxiety and the costs of additional testing). Due to this bias towards avoiding a miss, radiologists are taught to report any evidence that indicates the presence of a possible abnormality, even if this evidence is weak. Given that the knowledge reflected by the eye movement record on missed abnormalities was not reflected in the overt decision, it was likely that it was not consciously accessible at the time of the decision. Second, providing radiology experts with feedback about their eye movements has been shown to improve accuracy, suggesting that the eye movement record reflected information that was previously not consciously accessible to the expert (Kundel et al., 1990). As outlined below, although further work is required, this line of research has the potential to contribute to our understanding of implicit knowledge and expertise.

Given that radiology images do not have predefined interest areas, a set of procedures was developed for categorizing and grouping fixations. As detailed by Nodine et al. (1992), fixation clusters are formed by summing together nearby consecutive fixations. This is done by setting a spatial threshold (typically 2.5°) and then comparing the location of each new fixation to the mean location of all of the previous fixations in the current cluster. If the distance of the new fixation from the cluster mean is below a threshold of 2.5°, it is grouped with the previous fixations in that cluster, and if it exceeds the threshold, the new fixation marks the start of a new cluster. Given that regions are sometimes refixated, to compute the total time associated with a region throughout the trial (referred to as the cumulative cluster duration) clusters are summed whenever the centre of two or more clusters in a trial is closer than the threshold of 2.5°.

Cumulative clusters can be associated with four different response outcomes, depending on whether or not there was an abnormality in the fixated region, and whether or not an abnormality was reported. When the centre of a cumulative cluster is within 2.5° degrees of the centre of an abnormality, it is associated with a true-positive (TP) response (i.e. a hit) if the abnormality was later reported, and with a false-negative (FN) response (i.e. a miss) if the abnormality was not reported. For abnormality-free regions, cumulative clusters are associated with a false-positive (FP) response (i.e. a false alarm) if the observer had incorrectly reported an abnormality in a location that was less than 2.5° from the centre of the cumulative cluster. Cumulative clusters falling in unreported, abnormality-free regions are associated with true-negative (TN) responses (i.e. a correct rejection).

Subsequent research consistently documented that the longest cumulative cluster durations are associated with true-positive and false-positive responses, with somewhat lower durations for false-negative responses, and the lowest durations for true-negative responses. As can be seen from Fig. 29.5A,
the finding that cumulative cluster durations are longer for false-negative responses than true-negative responses is very general and was demonstrated for a variety of image types, including chest x-rays (e.g. Kundel et al., 1978, 1989, 1990; Krupinski and Roehrig, 2002; Manning et al., 2004, 2006b), mammography (e.g. Krupinski, 1996a, 1996b; Krupinski and Roehrig, 2000; Krupinski et al., 1999; Nodine et al., 2002; Nodine et al., 2001), bone trauma and fractures (e.g. Hu et al., 1994; Krupinski and Lund, 1997). As shown in Fig. 29.5B, this effect was also demonstrated across different levels of expertise (e.g. Krupinski 1996a, 1996b; Manning et al., 2006b).

The finding that eye movements consistently distinguish between false-negative and true-negative responses has been exploited by researchers in order to develop a visual feedback method for improving diagnostic accuracy (Kundel et al., 1990). The visual feedback method involves monitoring
radiologists’ eye movements while they make an initial decision about the presence or the absence of an abnormality, and then providing them with feedback about the locations and durations of their eye movements. On the basis of this feedback, the radiologist may then either confirm or revise their initial decision. Typically, the feedback is provided in the form of circles (5° in diameter) that are overlaid on top of the image to indicate the locations that were fixated for greater than a threshold duration that is typically set at 1000 ms (see Fig. 29.6A).

Although earlier work had discussed the possibility of using perceptual feedback about eye movements to improve performance (Nodine and Kundel, 1987), the idea was first tested by Kundel et al. (1990). Six radiology residents examined chest images for lung nodules for 15 s while their eye movements were monitored. After making an initial decision about the locations containing nodules, they either received feedback about the regions that they looked at for 1000 ms or greater, or they were given another look without any feedback. In both conditions, they were given an unlimited
amount of time during the second viewing, and they could either confirm or revise their initial decision. The same observers repeated the experiment 2 months later in order to counterbalance these two conditions. Importantly, in the feedback condition, performance improved by an average of 16% after the second view, relative to the initial view, while the no feedback condition led to an average decrease of 3% (see Fig. 29.6B). The performance benefit in the feedback condition reflected both an increase in the number of true-positives, and a decrease in the number of false-positives, and performance was measured by calculating the area under the AFROC curve (Chakraborty and Winter, 1990), which is a variant of the ROC analysis technique that is adapted for cases in which multiple responses are associated with a single image.

Building on the finding that visual feedback improves lung nodule detection, Barrett and Trainer (1997) suggested that local image enhancement (such as contrast adjustments) could be made in regions that received prolonged dwell times. Furthermore, it has been reported that both novice and expert radiographers show improved performance after viewing the eye tracking record of another observer who examined the same image, although the improvements were more pronounced for the novices than the experts (Litchfield et al., 2008). In addition to these eye tracking variations on the original method, researchers have proposed non-eye tracking feedback methods. These methods include the monitoring of interpretation times for mammograms (Saunders and Samei, 2006), and the pattern and duration of the zooming choices made by observers while using an interface that allows them to zoom in or out of particular regions (Mello-Thoms et al., 2000).

While the finding reported by Kundel et al. (1990) was replicated using another lung nodule detection task (Donovan et al., 2008), a 12% improvement in the case of mammography (Nodine et al., 2001) was not significant. In addition, for bone fractures feedback did not improve performance (Donovan et al., 2005). There are a number of factors that might influence the effectiveness of feedback, including the physical appearance of the cue itself (Krupinski et al., 1993a, 1993b), the experience of the observers and the complexity of the task (Donovan et al., 2005; Nodine et al., 2001), the amount of time spent viewing the image (Krupinski et al., 1998), and the threshold selected (Krupinski et al., 1998). As shown in Figs. 29.7 and 29.8, survival curves can be used to make predictions about the efficacy of different feedback thresholds (Krupinski et al., 1998). Survival curves are created by plotting the proportion of cumulative cluster durations that were at least as long as (i.e. survived) certain duration cut-off points. The four response types (TP, FP, FN, TN) are plotted separately in order to compare their different time-courses. For example, as can be seen in Fig. 29.7, for all image types, a greater proportion of FN responses than TN responses survived the 1000-ms cut-off point. For the purpose of feedback, a cut-off point should be selected such that it maximizes the number of circles containing abnormalities, and minimizes the number of circles without abnormalities. An inspection of the survival curves in Fig. 29.7 reveals that for the case of chest images the 1000-ms threshold is the most effective. In contrast, for mammograms, it appears that a higher threshold might be more effective than 1000 ms at differentiating between FN and TN responses. Finally, for bone trauma and bone fractures, although there is still a difference between FN and TN responses, the difference is very small, suggesting that feedback would be less effective for these image types, which is consistent with empirical findings (Donovan et al., 2005). In considering the effectiveness of the feedback, in addition to considering the different image domains and detection tasks, it might be important to consider the characteristics of the observer. The few studies that looked at the effects of expertise revealed that the FN and TN difference is quite consistent across levels of expertise. Krupinski (1996a) reported no significant differences between groups, although the difference between FN and TN cumulative cluster durations was numerically larger for radiologists than residents (see Fig. 29.5B). In addition, as shown in Fig. 29.8, Manning et al. (2006b) reported that FN responses were significantly higher than TN responses for all levels of experience. FP responses were longer than TP responses for the three more experienced groups, while the reverse was true for novices. Importantly, at least in the case of mammography and chest nodules, it appears that feedback could be effective for a wide range of levels of expertise.

To sum up, eye movements consistently distinguish between missed abnormalities and abnormality-free areas and it is possible that radiologists are not always aware of their prolonged fixations on
missed abnormalities. There is evidence that radiologists cannot accurately report the pattern of eye movements that they use to examine an image (Carmody et al., 1984) and due to the lenient response criterion adopted by radiologists (Scheff, 1963), the fact that they do not make use of all of the information reflected in the eye movement record suggests that at least part of this information was not available to them at the time of the overt decision. Consistent with this notion, providing radiologists with feedback about their eye movements has been shown to improve accuracy in the case of lung nodule detection (Kundel et al., 1990). However, although the feedback paradigm seems promising, further research is needed to determine the factors that maximize its effectiveness and to determine the mechanisms through which feedback improves performance. It has been suggested that feedback works by focusing attention on specific locations containing abnormalities while at the same time reducing interference from the surrounding regions (Krupinski et al., 1993a, 1993b). However, as discussed above, feedback might also be effective because it provides an additional source of information that was previously not consciously accessible to the radiologists.

### Conclusion

The present review illustrates that eye movement paradigms may prove invaluable in supplementing traditional measures of performance such as RT, accuracy, and verbal reports as a means for
Fig. 29.8 Survival curves (created by plotting the percentage of cumulative clusters that are longer than certain threshold values) as a function of expertise. Based on data from Manning et al. (2006b).

understanding human expertise in general, and visual expertise in chess and medicine in particular. Specifically, by employing eye-motion methodology, the research reviewed here provided powerful and direct evidence for the suggestion of de Groot (1946/1965) and Chase and Simon (1973a, 1973b) that a perceptual advantage is a fundamental component of chess skill and that in line with the global-focal search model (Nodine and Kundel, 1987) a global processing advantage is a crucial aspect of visual expertise in medicine. In addition, evidence derived from eye movement studies suggests the occurrence of unconscious or implicit processing of domain related patterns by experts, and such findings might begin to illuminate this important but as of yet relatively unexplored topic. In addition to its theoretical significance, the study of expertise and eye movements could lead to practical applications, such as more effective training programmes, expert systems and tools for computer aided diagnosis.

References


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


