Final author draft only. The copyedited article may differ from this manuscript version. The details of the article are as follows:

Donovan, T., & Litchfield D. (2013). Looking for cancer: Expertise related differences in searching and decision making. ***Applied Cognitive Psychology, 27,*** **43-49.**

Running Head: Visual Scanning of Lung Nodules

Looking for cancer: Expertise related differences in searching and decision making

Tim Donovan, PhD 1, 2 *

Damien Litchfield, PhD 1, 2

1 School of Medical Imaging Sciences,

University of Cumbria, UK

2 Department of Psychology

Lancaster University, UK

---

*Corresponding author:

School of Medical Imaging Sciences, University of Cumbria University, UK

Email: tim.donovan@cumbria.ac.uk

Tel: +44 (0)1524 384667

Abstract

We examined how the ability to detect lung nodules in chest x-ray inspection is reflected in experience-related differences in visual search and decision-making, and whether the eye-tracking metric time-to-first hit showed systematic decreases across expertise levels in medical image perception. Observers with a range of medical expertise undertook a free-response nodule detection task and decision-making improved with expertise, however, the time-to-first fixate a nodule showed only a non-significant trend to decrease with expertise. Surprisingly, naïve and expert observers both allocated less visual attention at nodule locations compared to 1st and 3rd year radiography students. This similarity in visual attention at nodules but not in decision-making was explained by the fact that naïve observers were more likely than experts to fixate and make errors on distracter regions. Time-to-first hit has been linked to expert performance in mammography, but in chest x-ray inspection, this metric was not sufficiently sensitive to demonstrate clear linear improvements across expertise groups. This brings into question the use of this stand-alone metric as an indirect measure of rapid initial holistic processing.

Keywords: radiology; expertise; eye tracking; nodule detection; time-to-first hit

Looking for cancer: Expertise related differences in searching and decision making.

About 40,000 people in the UK die every year from lung cancer, world wide it is the most common cancer with 1.61 million new cases diagnosed very year (Office for National Statistics, 2010). The first investigation for lung cancer is usually a chest x-ray. Chest x-rays are complex medical images in which anatomy is projected onto a 2D image with many overlapping shadows and patterns requiring a great deal of expertise to distinguish subtle abnormalities from normal features and come to a diagnostic decision about the image. Understanding the performance of experts in tasks such as chest x-ray interpretation is important because it will allow us to identify skills and strategies which can be taught. However, despite the expertise of radiologists errors are still a problem, with radiological error in the range of 2-20% for clinically significant or major error (Goddard, Leslie, Jones, Wakely, & Kabala, 2001).

Eye tracking methodology has been used to make significant contributions to understanding the type of errors made. For example, previous studies have demonstrated that when looking for pulmonary nodules, which can be precursors to lung cancer, false negative decisions (thinking a feature is normal when in reality it is a nodule) are made 30% of the time, but 70% of these missed nodules are, in fact, fixated or looked at (Kundel, Nodine, & Carmody, 1978; Manning, Ethell, & Donovan, 2004). This finding is also replicated in domains such as baggage screening (McCarley, Kramer, Wickens, Vidoni, & Boot, 2004).

One theory that can be used to explain the reproducibility of expert performance in radiology is the holistic model of image perception (Kundel, Nodine, Conant, & Weinstein, 2007) which proposes that experts perceive the image holistically, and that perturbations of the schematic normal anatomy are fixated and identified, if not perturbed, a focal search strategy is initiated, and this is followed by diagnostic decision-making. This model depends on the top-down global-level representations guiding processing of the details in an image, and so perception of pathology depends to a large extent on the ‘gist’ of an image obtained within the first 200ms (Carmody, Nodine & Kundel, 1981; Kundel & Nodine, 1975; Mugglestone, Gale, Cowley & Wilson, 1995). That is, prior to foveal search, experts are able to rapidly process the low-level information relating to the present image and contrast this information with their extensive experience of previously viewed normal and abnormal medical images. This enables experts to quickly recognise and coordinate their search for abnormalities much faster than less-experienced observers, without compromising on decision accuracy (Nodine, Mello-Thoms, Kundel, & Weinstein, 2002).

Since experienced radiologists identify abnormalities quicker and more accurately than less-experienced observers, the time it takes the eyes to first fixate the abnormality (i.e., ‘time-to-first hit’) can help discriminate between levels of expertise, and has recently been put forward as further supporting evidence for the holistic processing of medical images (Kundel et al., 2007; Nodine, Kundel, Lauver, Toto, 1996). For example, Kundel, Nodine, Krupinski & Mello-Thoms (2008) performed a mixture distribution analysis based on the time-to-first hit data from three mammography studies and found that over half of all cancers were fixated within one second, with the remaining cancers fixated in subsequent search. This provided compelling evidence of two discrete information processing systems involved in the detection of abnormalities; the rapid initial holistic processing that guides search, and the slower, subsequent processing relating to search and discovery.

These findings are broadly consistent with current psychological research regarding how other scenes and objects are processed (e.g., Castelhano & Henderson 2007; Henderson 2007; Schyns & Gosselin 2003; Wallis & Bulthoff, 1999; Wolfe, Võ, Evans, & Greene, 2011). Moreover, a recent meta-analysis by Gegenfurtner Lehtinan and Säljö (2011) looked at how the eye tracking data from professional domains supported the holistic model of image perception, Ericsson and Kintch’s (1995) theory of long term working memory and

Haider and Frensch's (1999) information-reduction hypothesis, and found that in line with the holistic model of image perception, experts from a range of domains tend to have faster time-to-first hits than novices.

But if time-to-first hit does decrease with expertise, is it necessarily the case that holistic processing is the underlying mechanism? For example, rapid serial search mechanisms[1] could also be exploited and account for the shorter latencies that accompany expertise**.** Indeed, if one examines the average fixation durations at normal and abnormal regions, fixation durations are typically short in normal regions, and in a way correspond to rapid serial true negative decisions, whereas fixations on suspicious areas lead to longer durations, and presumably extended cognitive processing (Manning et al., 2004).

Nevertheless, the studies that originally inspired the notion of rapid holistic processing operating in medical image perception were based on tachioscopic presentations that do not allow for eye movements (Carmody et al., 1981; Kundel & Nodine, 1975; Mugglestone et al., 1995), or by segmenting images and presented them serially, which disrupted experts' global processing advantage and reduced accuracy (Carmody, Nodine, & Kundel, 1980). Such direct manipulations of early scene viewing indicates holistic processing contributes to where observers look, in addition to subsequent serial search mechanisms. Whilst it is likely that both processes are running in parallel, the timings of such processes involve clear differences in perceptual and cognitive involvement (e.g., Wolfe et al., 2011). Moreover, the advantage in fixation guidance based on the holistic processing of the initial glimpse would likely diminish with time, and this has recently found to be the case; the initial glimpse facilitates up to the fourth fixation (Hillstrom, Schloley, Liversedge, & Benson 2012). Therefore, any supposed links of time-to-first hit with holistic processing would depend on the speed and difficulty of finding the abnormal regions within this time. Since Kundel et al. (2008) found

[1] We thank Tom Busey for bringing this to our attention

that over half of all cancers in mammography were fixated within one second such fixations may still be under the influence of the initial glimpse. However, given the complicated anatomical structure, the number of pathologies, and the heterogeneity of background distractors, chest x-ray inspection is one of the most difficult medical image perception tasks (Taylor, 2007). Therefore it is unclear whether time-to-first hit would fall within this range for this imaging modality.

Kundel et al. (2008) suggested that as expertise in medical image perception increases, it is more likely that recognition of abnormalities will be based on initial holistic processing. However, most medical image perception studies tend to compare intermediates (trainee registrars) and experts (radiologists), but not naïve observers (though see Kundel & La Follette 1975, Nodine et al., 1996). As such, we actually know remarkably little regarding the visual scanning, time-to-first hit, and holistic processing of naïve observers when viewing medical images. Therefore, if this initial holistic processing is so important in the transition from novice to expert, it seems prudent to encourage more studies that examine a broad range of radiological expertise.

Where direct comparisons have been made between experts and naïve observers, this has involved general visual search tasks (Nodine & Krupinski, 1998) and visual discrimination tasks (Sowden, Davies, & Roling, 2000), the outcomes of which point to the domain-specificity of expertise. This expert-focused approach to studying radiological expertise means we now have a better understanding of some of the mechanisms that engender expert performance. However, this can also give the impression that observer performance and efficiency of search improves in a linear fashion. For example, there is evidence to suggest that observer performance and efficiency of search can decline as the observer tries to apply newly gained knowledge (Lesgold et al., 1988; Wooding, Roberts, & Phillips-Hughes, 1998).

In previous studies (Manning, Barker-Mill, Donovan, & Crawford 2005; Manning et al., 2004; Manning, Ethell, Donovan, & Crawford, 2006) we have tried to take into account this need to understand medical image perception from a broad range of radiological expertise. We found that observer performance and overall search efficiency improved with expertise, and that different diagnostic decisions were reflected by different eye movement patterns. The present study builds on this approach of studying medical image perception and further examines the visual scanning behaviour of observers performing a nodule detection task. With the suggested links between time-to-first hit and holistic processing, our main goal was to establish whether time-to-first hit would show systematic decreases across expertise levels in medical perception, and particularly whether the time-to-first hit from experts would be fast enough to be under the influence of rapid initial holistic processing. We also explored how changes in eye movement behaviour reflect increases in diagnostic performance across the expertise levels.

Method

Participants. Forty observers with different levels of expertise took part in an eye tracking experiment in which they were required to search for lung nodules from a test bank of 30 chest images. The expertise groups consisted of 10 naïve observers, who were a mixture of students and staff from other disciplines within the university and had no experience of medical image perception, with a mean age of 29.2 years (range 22-40); 10 1st year undergraduate radiography students with 12 weeks clinical experience and a mean age of 31.6 years (range 19-48); 10 3rd year undergraduate radiography students with 28 weeks clinical experience and a mean age of 29.2 years (range 20-54); and 10 experts (8 consultant radiologists, 2 reporting radiographers) with a mean age of 47.2 years (range 35-55). Institutional ethics approval and informed consent was obtained from all participants.

Materials. Thirty images were selected from a previously compiled test bank consisting of natural nodules which were histopathologically proven and simulated nodules (see Manning et al., 2004). Half of the images contained nodules, some with multiple nodules (up to a maximum of five on any one image). Nodules were defined as discrete opacities in the lung field or mediastinum measuring between 5-30mm in diameter. Of the 28 nodules in total, 24 were histopathologically proven and the remaining 4 were simulated.

Eye-tracking data was acquired with a Tobii x50 system (Tobii Technology, Stockholm, Sweden), which samples eye movement position every 20ms. To determine the level of visual processing at nodule locations, rectangular regions of interest were defined around each nodule, the size of which was defined to be 1.5 times the diameter of the lesion. Fixations were defined as being within a 50 pixel radius for greater than 100ms (i.e., if a fixation point was more than 50 pixels away from the previous one it was classified as a separate fixation).

Observer performance was measured using a jackknife free-response ROC (JAFROC) analysis (Chakraborty, 2006). This generates a figure-of- merit that allows quantification of search performance; it is defined as the probability that an observer will rate a lesion higher than the highest rated non-lesion on a normal image. Unlike conventional ROC analysis which is only suitable for binary decisions (e.g., normal/abnormal), JAFROC analysis takes into account multiple decisions and/or abnormalities that occur on the same image and requires that observers specify the location of abnormalities, and as such, has greater statistical power (Chakraborty, 2000).

Procedure. Observers were instructed to look for lung nodules and disregard any other radiological findings. They were told that images may have no nodules, one nodule or multiple nodules. All observers were shown two chest images as practice examples, one normal and one with multiple nodules prior to the start of the study. Following a 9-point calibration, the first image was presented, free search was allowed, and observers were instructed to click on each nodule they perceived and indicate their confidence in their decision using a 1-4 rating system. Search was terminated when the observer pressed the space bar.

Results

True positive and false positive confidence ratings were subjected to a JAFROC analysis (Chakraborty & Berbuam, 2004). Observer performance was significantly different across expertise groups [$F(3,36) = 8.89$, $p < .001$]. Planned pairwise comparisons indicated that naïve observers performed significantly worse than all other observer groups (all *p*s < .05); however, there were no other significant differences across the groups. Although there was a clear trend for performance to improve with expertise (as evidenced by the higher figure-of merits in Table 1), this was overshadowed by inter-observer variability. For example, the JAFROC scores ranged from 0.07-0.74 for naïve observers; 0.37-0.82 for 1st year radiographers; 0.42-0.88 for 3rd year radiographers; and 0.45-0.81 for experts. Despite this interobserver variability in diagnostic performance, the total time taken to view each image was remarkably consistent across observer groups (Naïve: M = 31.63s; 1st year: M = 34.76; 3rd year: M = 35.96; Expert: M = 28.47), with neither group making final decisions faster than any other group [$F(3,36) = 0.60$, $p = .618$].

Eye movement measures at nodules were compared across groups and are reported in Table 1. Time-to-first hit was the time taken from when the image appears until a fixation is made in the ROI containing the nodule. In the case of images with multiple nodules, the minimum time to fixate any nodule was used. As shown in Table 1, average time-to-first hit for all expertise groups were in excess of 3 seconds; contrary to previous findings in mammography, 33% of cancers were fixated within 1s, whereas 56% (range 54-61% across expertise levels) of cancers were fixated within 2s, suggesting that pulmonary nodules are indeed harder to initially find than cancers in mammography. More surprisingly however was that time-to-first hit was not significantly difference across the expertise groups [$F(3,36) = 1.88$, $p = 0.14$]. There was a trend for time-to-first hit to reduce with expertise, particularly after the 1st year of training; however, observers were just as fast as each other at detecting nodules. Moreover, a Pearson's product correlation indicated that time-to-first hit was not significantly related to the JAFROC scores ($r = -.064$, $p = .696$).

The number of fixations made can be used be to indicate efficiency of search with the number of fixations overall negatively correlated with search efficiency (Rayner, 1998). Whilst there was no significant differences in time-to-first hit, there was a significant difference across expertise groups in the number of fixations at nodules [$F(3,36) = 5.41$, $p < 0.001$]. Posthoc comparisons revealed that experts made fewer fixations at nodules than 1st years ($p = .044$) but not 3rd years ($p = .084$), or naïve observers ($p = .93$). Similarly, naïve observers made fewer fixations at nodules than 1st years ($p = .007$) and 3rd years ($p = .017$), but there was no difference between 1st years and 3rd years ($p = .99$).

The cumulative dwell time of fixations in each ROI can also reveal the amount of visual and cognitive processing associated with abnormal areas. The experience-related differences in the number of fixations at nodules were mirrored in the dwell times [$F(3,36) = 5.91$, $p < .001$], with experts making shorter dwells on nodules than 1st years ($p = .001$) and 3rd years ($p = .025$), but not naïve observers ($p = .65$). Naïve observers made shorter dwells on nodules than 3rd years ($p = .037$) but not 1st years ($p = .34$), and there was no difference between 1st year radiographers and 3rd year radiographers ($p = .74$). Moreover, dwell time can be used to classify false negative errors relating to visual search, recognition and decision-making (Kundel et al., 1978). Using the convention[2] adopted by Mello-Thoms et al. (2005),

---

[2] Although earlier approaches discriminated between false negative errors based on individual samples (Kundel et al., 1978), Mello-Thoms et al. (2005) successfully used a

we classified false negative decisions into 1) visual search errors, where nodules were not fixated whatsoever, 2) recognition errors, where nodules were fixated for less than 1000ms, and 3) decision errors, where dwell times on nodules were greater than 1000ms. By reporting both the absolute and relative errors of each type in Table 2, it is clear that whilst naïve observers indeed make more visual search errors than other experience groups, they also make more recognition and decision-making errors overall. Thus, there does not appear to be a systematic decrease in error type with expertise.

To examine these issues further, for each group of observers we calculated the probability nodules were fixated, the probability of making a True Positive (TP) decision, and the probability of making a TP decision based on whether the relevant nodule was fixated. As shown in Table 3, there was no difference in the likelihood of fixating a nodule within a trial [$F(3,36) = 1.12$, $p = .356$]. There was however, a difference between the expertise groups in the likelihood of making a TP decision [$F(3,36) = 7.27$, $p < .001$], and the likelihood of making a TP decision based on whether the relevant nodule was fixated [$F(3,36) = 6.47$, $p < .001$]. Posthoc comparisons revealed that naïve observers were less likely to make a true positive decision than 3rd years ($p = .004$) and experts ($p = .003$). Similarly, having fixated a nodule, naïve observers were less likely to make a true positive decision than 3rd years ($p = .006$) and experts ($p = .007$). No other differences emerged between the groups (all *p*s > .105). Since total decision times were equal across groups, this suggests that it is not speed, but selectivity of information processing that better discriminates the performance between these experience groups.

1000ms threshold to discriminate qualitatively different false negative errors in mammography

Contrary to our predictions, we did not find that task-relevant areas (nodules) were fixated faster with expertise, but once fixated, such areas received different degrees of visual attention based on the level of expertise. Nevertheless, our results do not point to a linear change with expertise. Instead, the overall pattern of eye movement behaviour suggests that experts and naïve observers devoted less focal attention to nodules before making a decision when compared to 1st year and 3rd year radiographers. Critically, however, the JAFROC analysis suggests that experts were much more likely than naïve observers to make the correct decision using this economical allocation of attention. That is, the reduced attention at nodules by experts can be seen as efficiency in visual search and decision-making, but the same cannot be said for naïve observers.

One explanation for why naïve observers showed reduced attention at nodules is that greater attention may have been devoted to distracter regions that contain anatomy resembling the appearance of nodules, such as the mediastinum. As observed earlier (Kundel & La Follette, 1972), when compared to more experienced observers, naïve observers appear to spend a disproportionate amount of time fixating the mediastinum than the outer lung fields. An example of this distribution of search in our study is shown in Figure 1, whereby the relative dwell times of all fixations are collapsed for each observer group and plotted in a hotspot analysis. These data are from the same chest radiograph, with a subtle nodule in the apex of the right lung, and when comparing across the expertise groups it is clear that both the experts and naïve observers adopt a frugal approach, with attention devoted to a few key areas; for experts, focal attention is directed to the nodule, whereas for naïve observers, focal attention is directed to the hilar and mediastinum. In contrast, both the 1st year and 3rd year radiographers exhibit an exhaustive search that ensures that each aspect of the image has received extensive focal attention.

To formally assess this hypothesis that naïve observers were more likely to look at visually similar distracter regions, subsequent areas of interest were created around the lung fields and the mediastinum. Given the broad size of these areas we focused on general eye movement metrics that highlighted the distribution of search across these key areas. There was a significant difference in the percentage of fixations across the lung fields and the mediastinum [$F(1,36) = 44.68$, $p < .001$] and a significant area x expertise interaction [$F(3,36) = 6.63$, $p = .001$], however there was no main effect of expertise [$F(3,36) = 2.16$, $p = .111$]. To explore the significant interaction between expertise and area, a simple main effects analysis comparing across expertise levels at each area revealed a reliable difference for all expertise levels (all *p*s < .05) except for naïve observers ($p = .80$). Bonferroni-corrected pairwise comparisons confirmed that all observer groups except for naïve observers spent proportionally more time examining the outer lung fields rather than the mediastinum (see Table 1). Consequently, the results support the hypothesis that when searching for nodules, naïve observers are more likely to look at normal regions that contain visually similar distracters. More generally, these analyses clarify that although naïve observers and experts may show similar eye movement behaviour at nodules, these similarities can arise from different understandings of the medical image, and the ability to discriminate normal anatomy from pathology.

Discussion

Most expertise studies compare experts and novices and infer that any differences are the cause for the improved performance of experts. However, it is unusual for medical image perception studies to include naïve subjects or lay people. One exception is the early study by Kundel and La Follette (1972) which aimed to determine the evolution of the initial search strategy and fixation distribution from that of a layman to that of the skilled radiologist. Our study carried out 40 years later with a greater number of participants, more sophisticated eye tracking equipment, and 40 years of advancement in radiological training had very similar findings. JAFROC analysis indicated that observer performance improved with expertise, but there was substantial inter-observer variability. How long observers spend viewing images has been linked to performance, with increasing time spent viewing images actually leading to an increase in false positive decisions (Christensen et al., 1981; Edwards, Ricketts, Dubbins, Roobottom, & Wells 2003; Nodine et al., 2002). In our free-response task, all observers had equivalent decision times, and after 30 seconds on average of viewing time, the majority of nodules were fixated across all expertise groups. But even if most nodules were fixated, only experienced observers (3rd years and experts) were more likely to make the appropriate decision when compared to naïve observers.

A key goal of this study was to closely examine the initial visual and decision-making processes, and particularly how time-to-first hit changed across a range of expertise levels in chest x-ray inspection. Unlike eye tracking studies in mammography (Kundel et al., 2007; Nodine et al., 1996) time-to-first hit tended to decrease with expertise but this was not significantly different across expertise groups. The difficulty of obtaining large numbers of experts is often a problem in medical image perception studies, with typically 3-4 experts in each study (Kundel et al., 2008). Although our numbers of observers are much are higher than this, they are still low by psychological research standards, and this may have attributed to why we only found a downward trend in time-to-first hit across expertise levels.

Nevertheless, in our study it was evident that across observer groups 56% of cancers were fixated within 2s, whereas in mammography 57% of cancers were fixated within 1s (Kundel et al., 2008). This one second difference may not seem much, but given that the advantage of the initial glimpse diminishes quickly with time and each successive fixation (Hillstrom et al., 2012), this may in fact be crucial in highlighting the problem of associating time-to-first hit with mechanistic explanations involving holistic processing, and particularly in chest x-ray studies where abnormal regions are not immediately fixated.

Aside from these issues, our investigation revealed surprising commonalities between naïve observers and experts that were not shared between training radiographers and experts, for example, dwell times and the number of fixations at nodules were very similar for experts and naïve observers, when compared to radiographers in training. Thus, although observer performance appeared to improve in a linear manner with expertise, this was not the case for visual search as indicated by the eye tracking metrics. These similarities in attention at nodules can be explained when we take into account the overall distribution of visual scanning across the image, and specifically, the tendency for naïve observers to focus on areas of the chest radiograph with a low probability for containing nodules.

In terms of the performance of the 1st year and 3rd year radiographers, the pattern of results, although not significant, demonstrated the expected trend in observer performance and efficiency of search improving with expertise. Moreover, data from the first year radiography students indicated that after only 12 weeks clinical experience JAFROC scores improved and visual search was modified; with greater distribution of search towards the outer lung fields, and more attention directed at nodules. This pattern of improvement continued for 3rd year radiographers, with the significant exception that, when compared to experts, more attention was allocated to nodules before making a decision.

The current perspective on radiological expertise is that decisions regarding abnormalities are made early (Nodine et al., 1996; Manning et al., 2005) and that experts are better at making these fast and frugal decisions, presumably by exploiting initial holistic processing. It is generally thought that the expert has an efficient schema where the initial gist is matched with a high level, generalised idea of ‘normal’ achieved over many years of experience, whereas the naïve observer is limited by their representation of a nodule.

The goal of understanding radiological performance is to identify skills and strategies which can be taught. Previous attempts to train search strategies do not effectively tackle the problem (Carmody, Kundel, & Toto, 1984). Similarly, our efforts in a previous study to guide efficient search via examples of another's search behaviour (Litchfield, Ball, Donovan, Manning, & Crawford, 2010), though effective, are unlikely to be sustainable. Instead, there is a growing consensus that training in holistic processing may improve medical image perception (Donovan & Manning 2006; Kundel et al., 2008; Taylor 2007). However, this study highlights that as a stand-alone measure time-to-first hit is a problematic indicator of both holistic processing, and the development of expertise. If this eye movement behaviour is to be used to tap into holistic processing (and other possible mechanisms), more appropriate manipulations are needed. To that end, our future research is examining whether the 'flash-preview moving window' paradigm (Castelhano & Henderson, 2007) can be used to experimentally control how initial image understanding contributes to where observers look during medical image perception tasks, and how this develops with expertise.

Our broad approach to understanding radiological expertise has shown that with experience, there is not always a linear improvement in visual search and decision-making (cf. Lesgold et al., 1988; Wooding et al., 1998). More specifically, once training is underway, the transition towards expertise does appear linear (from 1st years, to 3rd years, and from 3rd years to experts), however, naïve observers with no formal instruction seem to exploit our effective visual search strategies and heuristics, that in some instances, have shown these observers to *look* like experts, even if they do not *decide* like experts. This suggests that future studies should consider these groups more carefully when trying to understand, explain and reproduce expert performance.

Acknowledgements

We are grateful to all the radiographers and radiologists who took part in this research.

References:

Carmody, D. P., Kundel, H. L., & Toto, L. C. (1984). Comparison scans while reading chest images: Taught but not practiced. *Investigative Radiology, 119*, 462-466.

Carmody, D. P., Nodine, C. F., & Kundel, H. L. (1980). Global and segmented search for lung nodules of different edge gradients. *Investigative Radiology, 15*, 224-233.

Carmody, D. P., Nodine, C. F., & Kundel, H. L. (1981). Finding lung nodules with and without comparative visual search. *Perception & Psychophysics, 29,* 594-598.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 753-763.

Chakraborty, D. P. (2006). Analysis of location specific observer performance data: Validated extensions of the jackknife free-response (JAFROC) method. *Academic Radiology, 13*, 1187-1193.

Chakraborty, D. P., & Berbaum, K. S. (2004). Observer studies involving detection and localization: Modeling, analysis and validation. *Medical Physics, 31*, 2313-2330.

Christensen, E. E., Murry, R. C., Holland, K., Reynolds, J., Landay, M. J., & Moore, J. G. (1981). The effect of search time on perception. *Radiology, 138*, 361-365.

Donovan, T., & Manning, D. J. (2006). Successful reporting by nonmedical practitioners, such as radiographers, will always be task-specific and limited in scope. *Radiography, 12,* 7-12.

Edwards, A. J., Ricketts, C., Dubbins, P. A., Roobottom, C. A., & Wells, I. P. (2003). The effect of reporting speed on plain film reporting errors. *Clinical Radiology, 58*, 971-979.

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

Gegenfurtner A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of Visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23,* 523-553.

Goddard, P., Leslie, A., Jones, A., Wakely, C., & Kabala, J. (2001). Error in radiology. *British Journal of Radiology, 74*, 949-951.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 172–190.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16*, 219-222.

Hillstrom, A. P., Schloley, H., Liversedge, S. P., Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychomic Bulletin & Review, 19,* 204-210.

Kundel, H. L., & La Follette, P. S (1972). Visual search patterns and experience with radiological images. *Radiology, 103,* 523-528.

Kundel, H. L., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*, 527-532.

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology, 242,* 396-402.

Kundel, H. L., Nodine, C. F., Krupinski, E. A., Mello-Thoms, C. (2008). Using gaze-tracking data and mixture distribution analysis to support a holistic model. *Academic Radiology, 15*, 881-886.

Lesgold, A. M., Rubinson, H., Feltovich, P. J., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In Chi MTH, Glaser R, eds. *The nature of expertise* (pp. 311-342). Hillsdale, NJ: Erlbaum.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied, 16*, 251-262.

Manning, D. J., Barker-Mill, S. C., Donovan, T., & Crawford, T. J. (2005). Time-dependent observer errors in pulmonary nodule detection. *British Journal of Radiology, 78,* 1-5.

Manning, D. J., Ethell, S. C., & Donovan, T. (2004). Detection or decision errors? Missed lung cancer from posteroanterior chest radiograph. *British Journal of Radiology,* 77, 231-235.

Manning, D. J., Ethell, S. C., Donovan, T., & Crawford, T. J. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules *Radiography, 12*, 134-142.

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual skills in airport-security screening. *Psychological Science*, *15*, 302-306.

Mello-Thoms C. Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C. et al. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Academic Radiology, 12*, 830-840.

Mugglestone, M. D., Gale, A. G., Cowley, H. C., & Wilson, A. R. M. (1995). Diagnostic performance on briefly presented mammographic images. *SPIE*, *2436*, 106-115.

Nodine, C. F, & Mello-Thoms, C. (2010). The role of expertise in radiologic image interpretation. In Samei E, Krupinski E, eds. *The handbook of medical image perception and techniques* (pp. 139-156). Cambridge: Cambridge University Press**.**

Nodine, C. F, Mello-Thoms, C, Kundel, H. L., & Weinstein, S. P. (2002). Time course of perception and decision making during mammographic interpretation. *American Journal of Roentgenology, 179*, 917-923.

Nodine, C. F., & Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Academic Radiology, 5*, 603-612.

Nodine, C. F., & Kundel, H. L. (1987) Using eye movements to study visual search and to improve tumor detection. *Radiographics,7*, 1241-1250.

Nodine, C. F., & Mello-Thoms, C. (2000). The nature of expertise in radiology. In J. Beutel, H. L. Kundel, & R. L. van Metter (Eds.), *The handbook of medical imaging: Physics and psychophysics* (pp. 859-894). Bellingham, WA: SPIE Press.

Nodine, C. F., Kundel, H. L., Lauver, S. C., & Toto, L. C. (1996). Nature of expertise in searching mammograms for breast masses. *Academic Radiology, 3,* 1000-1006.

Office for National Statistics (2010). Cancer statistics – registrations, England, 2008. Series MB1 no. 39. London.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Schyns, P. D., & Gosselin, F. (2003). Diagnostic use of scale information for componential and holistic recognition. In: Peterson MA, Rhodes G, eds. *Perception of faces, objects and scenes* (pp. 120-145). Oxford: Oxford University Press.

Sowden, P. T., Davies, I. R., & Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance, 26*, 379-390.

Taylor, P. M. (2007). A review of research into the development of radiologic expertise: Implications for computer-based training. *Academic Radiology*, *14*, 1252-1263.

Wallis, G., & Bulthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Science, 3*, 22-31.

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Sciences, 15, 77-84.*

Wooding, D. S., Roberts, G. M, & Phillips-Hughes J. (1998). The development of the eye movement response in the trainee radiologist. *SPIE, 3663*, 136-145.

Table 1.

The mean (and standard error) observer performance and eye movement behaviour across the four expertise groups.

Observer Performance		**Fixations at nodules**			**Distribution of search**	
Level of Expertise	JAFROC (figure-of merit)	Time-to-first-hit (s)	Number of fixations	Dwell time (s)	Lung Field %	Mediastinum %
Naïve observers	0.41 (0.06)	3.83 (0.53)	6.07 (0.88)	3.81 (0.56)	0.40 (0.04)	0.42 (0.06)
1st years	0.60 (0.04)	4.11 (0.70)	7.85 (1.44)	4.15 (0.56)	0.56 (0.05)	0.23 (0.04)
3rd years	0.71 (0.03)	3.27 (0.40)	7.31 (0.71)	4.33 (0.40)	0.51 (0.02)	0.24 (0.02)
Experts	0.72 (0.02)	3.31 (0.58)	5.69 (0.44)	3.13 (0.36)	0.60 (0.03)	0.20 (0.02)

Table 2.

Absolute (and relative) false negative errors classified into the number of visual search, recognition, and decision-making errors for each expertise group.

Level of Expertise	Visual search errors	Recognition Errors	Decision-making Errors
Naïve observers	34 (22%)	51 (32%)	72 (46%)
1st years	18 (18%)	40 (39%)	44 (43%)
3rd years	13 (17%)	23 (29%)	42 (54%)
Experts	13 (17%)	37 (48%)	27 (35%)

Table 3.

The probability for each expertise group to a) fixate the nodules, b) make a true positive decision, and c) make a true positive decision having fixated the relevant nodule.

Level of Expertise	Probability nodule fixated	Probability of true positive	Probability of true positive having fixated relevant nodule
Naïve observers	0.87	0.44	0.40
1st years	0.91	0.64	0.58
3rd years	0.93	0.72	0.67
Experts	0.92	0.73	0.67

Figure 1.

A hotspot analysis of visual search behaviour across the four expertise groups. Greater intensity represents longer dwell times at fixated location. A lung nodule is located in the upper right lung (top-left of the image).