A Review of Research into the Development of Radiological Expertise: Implications for Computer-based Training

Paul Taylor

University College London

Abstract

Studies of radiological error reveal high levels of variation between radiologists. Although it is known that experts outperform novices, we have only limited knowledge about radiological expertise and how it is acquired. This review identifies three areas of research: studies of the impact of experience and related factors on the accuracy of decision-making; studies of the organisation of expert knowledge; and studies of radiologists' perceptual processes.

Interpreting evidence from these three paradigms in the light of recent research into perceptual learning and studies of the visual pathway has a number of conclusions for the training of radiologists, particularly for the design of computer-based learning programs which are able to illustrate the similarities and differences between diagnoses, to give access to large numbers of cases and to help identify weaknesses in the way trainees build up a global representation from fixated regions.

Introduction

In 1999 the Institute of Medicine's Committee on Quality of Health Care in America produced a report 'To Err is Human' which contained the headline-grabbing statistic that as many as 98,000 Americans might be dying each year as a result of medical error [1]. The report was highly influential and led professional organisations and government agencies, not just in the United States, to pay increased attention to the problem of medical error and the concept of patient safety.

Most studies of error in radiology report significant rates of 'observer variability' or disagreement with a gold standard. A review by Goddard et al. suggests a range of 2-20% for clinically significant or major error across radiological investigations [2]. More recently van Rijn et al. compared spiral CT with MRI in a series of patients suspected of lumbar herniated discs and found only moderate levels of agreement between observers, who, for example, disagreed on herniation at CT evaluation in 12% of discs [3]. Halligan et al. asked experts, consultants, and experienced trainees to report a consecutive series of 20 double-contrast barium enema studies [4]. They found that experts misclassified 23% of cases whereas consultants misclassified 31% and trainees 34%. Manning, Ethell and Donovan report a missed lesion rate of 27.2% in a study involving 4 radiologists viewing 120 chest X-rays, containing 81 lung nodules [5]. Monnier-Cholley carried out an ROC study of lung cancer detection from a selected set of chest X-rays and noted that overall consistency in observer detection of lung nodules was poor (the LROC A_z being 0.54 $\pm$

0.024 for residents, 0.561 ± 0.033 for staff radiologists) [6]. These findings should be considered in the light of others suggesting that clinically insignificant errors predominate over significant ones and that audits of large numbers routine cases reveal lower error rates than laboratory studies based on selected cases [7,8].

Many possible solutions to the problem have been proposed, including the use of computer aids, multiple human readers and improvements in working conditions [9,10,11]. The evidence about the effectiveness of computer aids is increasingly negative [12], constraints on manpower limit the scope for double reading while there is such variety in working conditions that general suggestions for improvement are not helpful. Another key variable which needs to be addressed is that of radiologist skill or expertise. We know that some radiologists are better than others, and that experts do better than novices. But what do we know about the nature of this difference? And does what we know help us to understand how training can best enhance radiologists' abilities? In the rest of this article I first review the research into radiological expertise and then summarise some more general findings about perception and experience. In the conclusion the reviewed research on expertise is interpreted in the light of the more general work in order to furnish some conclusions for radiological training.

The science of radiological expertise

Studies of radiological expertise can be considered under three headings: work on decision making, on the representation of expert knowledge and on perceptual processes.

Studies of medical decision-making

The concept of expertise is closely associated with a capacity for accurate decision-making. Much psychological research has been directed at understanding decision-making, and many of these researchers have considered the clinical domain. One paradigm in the study of experts' decision making uses a mathematical technique known as judgement analysis to identify the relative weights assigned to different diagnostic cues (see Wigton for a review of this work applied to clinical decision making [13]). Another approach involves comparing human decision making to theoretically optimal decision-making (maximising a measure of expected utility), often in order to identify 'cognitive biases' [14]. Interesting lessons have been learnt, judgement analysis, for example, reveals that decision-makers rely, for the most part, on small numbers of cues – experts making complex decisions rarely make significant use of more than four cues [15]. Cognitive biases identified in clinical decision-making include being over-influenced by information from recent or highly salient events – for example doctors who have recently encountered a rare condition might be over-willing to consider it as a diagnosis – and under-using information about prevalence [16].

Research in radiology often uses radiologists' decision-making as the dependent variable, but generally it is a radiological technique that is being investigated rather than the

decision-maker. Some studies have, however, looked at how decision-making is affected by important variables such as the availability of clinical information, or the experience of the radiologist. Swensson and colleagues carried out a series of studies to provide empirical support for the idea that searching in the absence of preconceptions allows for more accurate judgements about images. Swensson, Hessel and Herman 1985, for example showed that residents did better at distinguishing normal and abnormal features when instructed to search films and report all abnormalities that when primed to search for nodules or when asked to look at specified locations and say if a nodule was present [17].

This effect seems to be due to quite subtle characteristics of the task. Berbaum, Franken, Dorfman et al., for example, asked residents to read chest X-rays with and without additional clinical information, in the form of a prompt that was correct for abnormal cases and plausible for normals. ROC analysis showed an improvement in diagnostic accuracy for diverse subtle lesions, although not for simple nodules [18]. The effect seemed to be perceptual rather than decision-making since the effect was not explained by a reduction in false positives but rather involved the improved detection of lesions. More recently a systematic review by Loy and Irwig found 39 articles studying the impact of clinical information on the accuracy with which diagnostic tests were interpreted [19]. They reviewed the results of 16 studies, including five dealing specifically with chest X-rays and three with bone radiographs. Overall they found that the use of clinical information led to improved test reading accuracy, although there was

wide variation in the scale of the effect. They also suggest that the improvement seems to come more from improved sensitivity than specificity.

There are different aspects of experience that can be tested to investigate how they affect performance. At least one study has looked at the impact of a training course and found that it improved diagnostic accuracy [20]. A number of studies of expertise compare readers at different levels of qualification and show that staff in more senior positions are more accurate [see e.g. 21,22]. Studies which look at experience either in terms of years of experience or the number of cases viewed tell a less consistent story, at least as regards screening mammography. Beam, Conant and Sickles found a negative correlation between reading performance and years since qualification [23]. Barlow et al. report that more recently qualified radiologists have higher sensitivity but lower specificity which might suggest that years of experience lead to a change in the threshold for recalling a patient rather than improved accuracy [24]. However Smith-Bindeman et al. have found that radiologists with more years of qualification had better specificity and slightly improved or at least equal sensitivity, suggesting that the improved specificity is the result of more accurate interpretation [25]. Their analysis is summarised in Figure 1. Esserman et al. found that radiologists reading more films achieved greater specificity for a given sensitivity. [26]. Beam, Conant and Sickles, however, found no relationship between accuracy and the volume of films being read. They argue that expertise is a complex phenomenon and hypothesise that the quality of a radiologist's experience determines how it impacts on performance.

As well as these investigations into the impact of clinical information or experience on decision-making, there have been studies that aim simply to assess the variability in the accuracy of radiologists' interpretations, perhaps most notably in mammography. Wagner, Beam and Beiden. calculated the impact of radiologist variability on utility functions plotted in ROC space [27]. They argued that debate about the value of mammography, and, for example, about the potential of computer aids, must be informed by an understanding of the limited usefulness of poorly performing radiologists.

Studies of the organisation and structure of radiological knowledge

Research into the role of expertise in clinical reasoning has focussed, in recent years, on attempts to understand how an expert's knowledge is represented. One of the central ideas in this latter work is that of the schema, or script [28,29]. The notion of schema is applied by psychologists to many areas of cognition, the central idea being that human reasoning, for the most part, employs knowledge embodied in semi-automated plans for dealing with the routine parts of daily life. A common approach in this research is to analyse statements recorded when experts are asked to think aloud while solving complex problems.

Lesgold and colleagues carried out a classic series of such studies examining the nature of radiological expertise [30]. They compared the 'think out loud' reports of experts (radiologists with at least 10 years' of experience post-residency) with those of residents

at different stages in their training, in order to reveal how the acquisition of expertise alters the interpretation of radiological images. The first step in the analysis was to count the number of 'findings' – specific properties attributed to the film or the patient – each radiologist made. The researchers also counted how many times each radiologist made a connection between two findings. Experts identified more findings than residents and these findings were less likely to be unconnected observations, and were generally organised into clusters of connected observations (this, perhaps, explains why experts' decision rules, modelled in judgement analysis, are so parsimonious). Experts also performed more inferential thinking: longer chains of reasoning could be identified from the transcripts of their reasoning.

Lesgold et al. suggest that radiological schemata include a set of processes which will generally lead to a successful interpretation for a given class of cases, associated with a number of tests or prerequisites to ensure that the appropriate schema is selected. They argued that expert radiologists can, very efficiently, identify the best schema to guide the interpretation of the case; that they apply schema early, within the first two seconds, but continually test and elaborate the applied schema and can be flexible about changing schemata during interpretation.

These conclusions could however be questioned by anyone inclined to be sceptical of the method. The transcription and analysis of the transcripts is time-consuming and studies tend to be limited in terms of the numbers of subjects and images involved. There are also problems with the methodology: thinking aloud is not a natural part of reasoning.

Norman argues that at least some of the results obtained using this technique are probably artefacts of the method [31].

Norman reviews research exploring other hypotheses about the representation of clinical expertise including formal knowledge of basic science, decision trees, mental probability matrices or what are termed semantic qualifiers (networks of simple Boolean variables – e,g, acute vs chronic pain, large vs small joint) and notes that evidence has been found to support each [32]. He suggests that experts probably have access to knowledge represented in different forms. Norman's own work has considered two theories of how knowledge of diagnostic categories is represented. One 'the prototypical view', holds that category membership is determined by a set of explicit criteria, perhaps embodied as features of a prototype for that category; the other, 'the exemplar' view, considers that our mental categories are defined by the set of remembered category members so that membership is assigned following a search for a sufficiently similar exemplar. Although he and his colleagues have not considered radiology directly, they have addressed other visual tasks including dermatological diagnosis and ECG interpretation. In one experiment pre-clinical medical students studied a series of dermatological slides covering different conditions and were later tested on new slides [33]. The new slides included some that were carefully chosen to be either typical of the categories represented in the original slides or similar to the actual slides, but not both. Results showed that students were better if the new slides were typical (as the prototype theory would predict) but also better if the new slides were similar to the earlier ones (as the exemplar theory would predict). A similar result was found by Ark, Brooks and Eva who

compared the effects of priming students to assign a diagnosis on the basis of prototypical features or on the basis of an impression of familiarity or both and found that students did best when using a combined strategy [34].

However the value of a simple feature-based approach is challenged by a striking study reported by Kulatunga-Moruzi, Brooks and Norman. Participants (a mix of experienced dermatologists, family medicine specialists and residents) made dermatological assessments under three conditions: from a photograph, from a comprehensive list of the clinical features on the photograph – these lists were worked up by experts – and a combined condition in which the participants first read the comprehensive list then viewed the photograph [35]. Their performance was better in the first than the second condition and also better in the first than in the third condition. It is perhaps unsurprising that it is easier to make a diagnosis from an image than from its description, it is striking however that the extra information we get from our own visual processing is so important that being primed with an accurate and comprehensive verbal account, based on someone else's processing, actually damages our capacity to interpret the case correctly. The finding is similar to that of Swensson, Hessel and Herman, above [17]. Understanding radiological expertise requires an understanding of radiologists' perceptual processes.

Studies of radiologists' perceptual processes

An alternative source of data about interpretation is obtained by monitoring eye movements. It is possible, using widely available equipment, to track a radiologist's gaze and monitor which regions of the images are looked at and for how long (this work is well reviewed in [36] and [37]). This can tell us something about why abnormalities are missed, for example. In a classic early study of the eye movements of radiologists, in this case looking at chest X-rays, Kundel, Nodine and Carmody (1978) classified missed cancers according to whether the radiologists' gaze completely missed the abnormality (errors of search), passed over it without dwelling on it long enough for an object to be recognised (errors of detection) or dwelt on it but did not correctly identify it (errors of decision); 30% of missed nodules correspond to errors of search, 25% to errors of detection and 45% to errors of decision [38]. Patterns of search show a clear evolution in the course of a radiologist's training, becoming much more efficient, with the most dramatic changes occurring in medical school, apparently more affected by improvements in knowledge of radiographic anatomy and clinical medicine than by formal training in radiology [39]. Although errors of search are significant, a substantial majority of errors occur in regions of the image that appear to receive visual attention. Nearly half of those in the Kundel, Nodine and Carmody study occurred in regions that received prolonged attention. Mello-Thoms, Hardesty, Sumkin et al., classified errors in mammograms as Kundel, Nodine and Carmody had done, finding that 30% of missed lesions corresponded to errors of search, 40% to errors of detection and 30% to errors of decision [40]. They also collected data using images on which a visible cancer had been missed at the original screening exam. In these cases 47% of missed lesions corresponded to errors of search, 33% to errors of detection and 22% to errors of decision.

The benefits of experience come into play very early in the processing of an image. Kundel and Nodine found that experienced radiologists could identify 70% of the lesions on chest X-rays when presented with the films for just 200ms [41]. Mugglestone, Gale, Cowley and Wilson repeated the experiment with experts reading a test set of mammograms, radiologists correctly identified 51% of lesions in the 'flash' condition compared to 69% when given unlimited time to view the images [42].

Christensen, Murray, Holland et al. compared the time faculty members and residents took to record observations on 100 selected chest radiographs. For experts the curves for true positives tend to be convex, with most observations being made very quickly, while those for false positives increase at a linear rate [43]. With residents the true positive curves level off more slowly and the false positive curves begin to curve upwards at the end of viewing time, sometimes crossing those for true positives. Plotting observations per unit time on a log scale against time produces a curve that can be modelled as the sum of two straight lines, suggesting that the results are produced by two distinct processes, a rapid, almost instantaneous recognition of obvious abnormalities, followed by a slower search of the image. Nodine CF, Mello-Thoms C, Kundel HL, et al. found a similar effect plotting positive predictive value against time to observation for mammographers and trainees [44]. Nodine, Kundel, Mello-Thoms et al. analysed the accuracy of mammographic interpretation of three groups: technologists, radiology residents and staff mammographers with at least five years experience [21]. They too found that staff radiologists were both faster and more accurate than residents. In all these studies

decision times tend to be longer for false compared to true outcomes. Comparing the time course of decision-making for groups at different levels of expertise shows the overall accuracy of less expert groups tailing off as false-positive results begin to outnumber true positive results.

A recent series of experiments by Mello-Thoms and colleagues has used image processing to analyse fixated regions on mammograms. The technique is inspired by perceptual psychology which suggests that the eye responds to the spatial frequencies in an image. (Spatial frequency is the distance between elements in a repeating pattern and an indication of the scale at which useful information is detected. High spatial frequencies correspond to fine detail, lower spatial frequencies to coarser elements.) Mello-Thoms, Dunn, Nodine and Kundel measured the log of the energy (energy being the integral of the signal strength) of different spatial frequency bands across a range of orientations, for each region fixated by a radiologist [45]. The profile of the log-energy measurements is obtained for each fixated region, this profile contains the information that we believe is computed by the low level visual system. Since a radiologist makes comparative judgements, the experimenters consider not just the profile in each region (termed the local profile) but also a combined profile (the global profile) across all fixated regions in an image. The local profile is related to the conspicuity of local features, the global profile is a measure, specific to each observer, of the searched background to which local features are compared. Figure 2 shows the scan paths of two radiologists in the study and clearly demonstrates that the two are inspecting different areas on the film.

Using this technique to compare the decision-making of full-time mammographers with that of residents, Mello-Thoms found that the computed profiles did not discriminate between mammographers' true and false positive decisions [22]. This suggests that mammographers are making consistent use of a representation of malignant appearance which is well modelled by the spatial frequency representation. The profiles did, however, discriminate between mammographer's true positive and false negative decision outcomes. Conversely for residents, true positives *could* be distinguished from false positives but not from missed cancers. Interestingly the residents in this study principally made perceptual rather than decision errors and the global profile seemed to play a much smaller role in their decision-making. Mello-Thoms concludes mammographers are better able to contrast local findings with global features, while residents are more guided by the local conspicuity of image features. The most striking result is that an artificial neural network (a form of software which when trained on examples can learn to classify complex data items) was able to use the profiles to predict how a radiologist would report an image region. The mammographers, as one might expect, behaved more predictably than the residents.

Understanding the nature of perceptual learning

Recent years have seen a great deal of research into perceptual expertise or perceptual learning, and psychologists and neurobiologists have combined to develop models of how the performance of visual tasks is improved by practice. A convenient framework for

reviewing this research is a computational model of visual object recognition in which perceptual processing computes a hierarchy of representations starting with processing of the retinal image to extract information about intensity changes across a range of orientations and spatial frequencies, this information is then processed at higher levels to compute successively more complex representations, eventually arriving at something which can be matched against the stored representations that make up our visual memory (see [46] for a classic account of this approach to understanding vision). Perceptual learning occurs when the capacity of a subject to recognise an object is enhanced by practice. This enhancement must reflect a change in the visual pathway in how low or high level representations are computed or processed. Research by Ahissar and Hochstein attempted to identify the level at which the change occurs, using the intuition that learning at the higher levels of visual processing would transfer across different tasks, whereas learning at the lower levels would be specific to low-level features of the stimuli used [47]. They found that where the learning task was made more difficult, there was evidence of low-level learning, but that this only occurred after higher level learning had taken place. Their findings led them to develop the Reverse Hierarchy Theory of perceptual learning, shown in Figure 3: changes in high level representations guide the selective tuning of low level perceptual processing, in conditions when improvements in signal-to-noise processing are required.

Sowden, Davies and Roling carried out a series of studies to demonstrate that these kinds of improvements in visual sensitivity are involved in the acquisition of radiological expertise [48]. They found radiologists were better than non-radiologists at detecting

small low contrast dots placed at random against a background of nested rectilinear structures, an artificial detection task for which radiologists' conceptual knowledge and decision making skills would be irrelevant, suggesting that they had developed heightened sensitivity to these kinds of stimuli. In a further experiment they identified that a rapid increase in the performance of novice film readers could be obtained as a result of a few days experience with this kind of task, and that the improvement did not transfer to images in which the contrast was reversed (as it is in a photographic negative) suggesting that the improvement is due to low-level sensory learning.

In the future we may be able to draw on quite new sources of evidence to answer these questions. Haller and Radue carried out fMRI scans of 12 experienced radiologists and 12 non-radiologists performing a perceptual processing task on radiological and control images [49]. They found that the radiological images excited regions in the radiologists' brains which were not activated in those of non-radiologists. The authors note that activations in these regions (bilateral middle and inferior temporal gyrus, bilateral medial and middle frontal gyrus, and left superior and inferior frontal gyrus) have previously been associated with visual attention and memory retrieval, suggesting that the radiological images automatically attract attention and invoke memories of reference images. Interestingly the control images (electron microscopy images) also led to different patterns of activity between the two groups, including activations which have associated with the generation of a mental representation of an object and the mental rotation of objects in two- and three-dimensional space. The authors speculate that radiologists may create a mental representation of the presented image, because such

images are sometimes rotated and the position of anatomic structures may vary, they develop an enhanced capacity mentally to rotate the presented objects.

The relationship between cognition and perception in radiology

It is clear that the interpretation of radiological images involves the processing of characteristic perceptual data and that interpretive skill is in part a matter of skilled perception. The task is, however, not a purely perceptual one. We know, for example, that expert radiologists make more accurate assessments in the light of additional clinical information about a case [19]. One might think of the cognitive element of the task as being more reflective, involving conscious thought rather than the automated processing of visual information. The two elements are not easily separated however.

The advantage experienced radiologists have over their junior colleagues seems to come into play very early in processing and to involve the processes which guide the collection rather than the assessment of evidence. Recent research using magnetoencephalography and functional MRI to measure activity in the visual cortex suggests that early analysis of low spatial frequencies triggers activity in the left orbitofrontal cortex, which takes place concurrently with the slower propagation along the visual pathway of both low and high spatial frequencies. Bar, Kassam, Ghuman et al. hypothesise that this activity serves as top-down facilitation of recognition: the rapid analysis of coarse-grained information

serving to generate predictions about what might be in the image, which are then matched with the results of processing higher spatial frequencies [50].

The notion of attention, which is to do with the allocation of processing capacity or a selective emphasis on certain elements of information, is a cognitive rather than a strictly perceptual one. Selective attention is a way of optimising the processing of information for a particular purpose, e.g. searching for a particular target. It modulates our response to information gathered, including information from peripheral vision. This view of attention is supported by recent research which suggests that attention influences visual information even in the earliest steps in visual processing [51]. One theory, supported by evidence from neurobiology, is that the visual system creates an "integrated saliency map", which combines information about the 'conspicuity' of visual features and an assessment of their behavioural relevance, combining sensory information from low-level processing with high-level knowledge about what should receive attention [52]. The salience of a feature on a radiological image might then be determined, in part, by the conspicuity of the feature, and, in part, by prior knowledge and expectations. Evidence from studies of neuronal activity suggests that an assessment of salience guides eye movements [53].

Radiological expertise has a number of low level components. These are responsible for the kinds of effects observed by Sowden, Davies and Roling - the selective tuning of elements in early visual processing - and the computation of a representation of salient regions of mammograms in terms of spatial frequency profiles. It also seems to involve

something that resembles the cognitive scientists' concept of schemata: representations of habitual behaviours. The experienced radiologists studied by Lesgold et al. continually tested the appropriateness of the activated schema, monitoring the activity of the schema-directed search and the perceptual processing it leads to, and monitoring their own reaction to it. Research reviewed above on the time-course of decision-making shows that, for both experts and non-experts, the longer the search time, the more positive predictive value lowers off as false positive responses begin to predominate [43,44]. In real life radiologists must decide when to stop searching an image. This decision is based on a judgement not just about the image but about the radiologists' confidence in the accuracy of their assessment of the image. The radiologist must use a subjective response, a feeling of confidence, as the basis for a decision. They are reasoning that, after a certain point, had there been something in the image, it would have attracted their attention by now.

Implications for radiological training

A recent review concluded that there was very little research into the question of how radiologists are trained [54]. The research reviewed above, however, does suggest a number of tentative conclusions for the design of training programmes, and particularly of the scope for computer-based learning.

(1) If we assume that object recognition is primarily a matter of matching the output of low level processing with stored exemplars of known categories, it follows that the

primary goal of early training should be to provide radiologists with an appropriate set of exemplars. Some interesting work on computer-based learning has provided students of MRI with a tool which arranges examples from a training database in a conceptual space [55]. Each case is classified on a finite set of dimensions – each feature recorded about a case is represented as a value along a scale of some kind. The total set of cases can then be considered as a set of points in a multi-dimensional space. The tool displays a variety of two dimensional projections of the space. This can help trainees to get a sense of the way in which a set of instances of a disease are distributed around a mathematically 'typical' case and to assess the 'distance' between a new case, of unknown diagnosis, and the typical presentations of candidate diagnoses.

For example, in Figure 4, let the squares be the cases in the database of Syndrome A and the triangles images be those of Syndrome B, and the diamond a new image of unknown diagnosis. The idea is that the display helps the user pick out images from the database with which to compare the new image, in order to arrive at a classification. The centres of the circles, and the arrangements of the cases in the 'space' represent statistical measures of the typicality and similarity of the cases and are derived from the image findings to maximise the information provided by the display.

(2) Sowden, Davies and Roling experiments support the idea, suggested by Ahissar and Hochstein's Reverse Hierarchy, that expertise requires the tuning of low-level perceptual elements. Training programmes should be designed to support first the acquisition of appropriate knowledge then to allow the required perceptual learning to take place. Some

aspects of experience seem to have a greater impact on performance than others. If it were simply a matter of the number of cases read, the more cases a radiologist saw, the better they would be. This is not exactly what we find. The effect of years of experience is much stronger than that of number of cases read. It might be that radiologists who view more cases per annum are not being given sufficient feedback on performance, or adequate opportunity for reflection for the experience to be beneficial. Smith-Bindeman et al. report radiologists reading 2500-4000 mammograms a year had improved specificity but that this effect was not found in radiologists reading more than 4000 per annum [25]. We need to ensure that all radiologists, but particularly recently qualified radiologists, are given not just experience but also the opportunity to learn from it.

There is therefore an argument for building large databases of cases to be used for training. Ericsson, writing about expertise more generally, has argued that it requires a form of 'deliberative practice' and that the scarcity of outstanding performance is in large part because few people have the opportunity to engage in this [56]. A number of collaborations are now putting together substantial collections of annotated digital cases [57, 58]. Accessing these databases through appropriate tools may allow trainees to acquire the benefits of experience more efficiently than they can through on the job learning.

(3) Students are frequently told to scan images systematically, and yet few radiologists seem to work this way [36]. The impact of training, is not that readers learn the right way to move their eyes across the image but that training provides a cognitive basis for

recognising 'perturbations', deviations from the radiologist's mental model of normal appearance. Kundel, Nodine, Conant and Weinstein suggest that experts can analyze the visual input of the entire retinal image holistically and locate an abnormality on the basis of this global processing [59]. Analysis of the spatial frequencies in fixated regions appears to tell us something about the basis of radiological decision-making and highlights the role of search in building up an accurate representation of the image background, in order to discriminate between true and false positives. Providing residents with clear feedback about their failure in this aspect of radiological search could lead to a rapid improvement in performance. Mello-Thoms has shown that an artificial neural network trained on this data can anticipate radiologists' decisions [45]. Such a tool might have a role in computer-based training, as a model of the representation used by the radiologist to guide their decision-making.

Conclusions

The development and acquisition of radiological expertise has been studied for thirty years. Three distinct paradigms can be identified: studies of the impact of various factors, including experience, on the accuracy of radiologist decision-making; studies of the organisation of expert knowledge; and studies of perceptual processes. Although there has been relatively little research into the effectiveness of different approaches to training radiologists, the findings reviewed here suggest some interesting ways in which training interventions, particularly computer-based learning, could be designed.

References

1. Kohn LT, Corrigan JM, Donaldson MS. Executive Summary in *To Err is Human: Building a Safer Health System* Institute of Medicine, National Academy Press, Washington DC 2000.

2. Goddard P, Leslie A, Jones A, et al. Error in radiology. Br J Radiol. 2001 Oct; 74 (886):949-51

3. van Rijn JC, Klemetso N, Reitsma JB et al. Observer variation in the evaluation of lumbar herniated discs and root compression: spiral CT compared with MRI. Br J Radiol. 2006 May;79(941):372-7.

4. Halligan S, Marshall M, Taylor S, et al. Observer variation in the detection of colorectal neoplasia on double-contrast barium enema: implications for colorectal cancer screening and training. Clin Radiol. 2003 Dec;58(12):948-54

5. Manning DJ, Ethell SC, Donovan T. Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph. Br J Radiol. 2004 Mar;77(915):231-5.

6. Monnier-Cholley L, Carrat F, Cholley BP, Tubiana JM, Arrive L. Detection of lung cancer on radiographs: receiver operating characteristic analyses of

radiologists', pulmonologists', and anesthesiologists' performance. Radiology. 2004 Dec;233(3):799-805

7. Erly WK, Berger WG, Krupinski E, et al. Radiology resident evaluation of head CT scan orders in the emergency department. Am J Neuroradiol. 2002 Jan;23(1):103-7

8. Lal NR, Murray UM, Eldevik OP, Desmond JS. Clinical consequences of misinterpretations of neuroradiologic CT scans by on-call radiology residents. Am J Neuroradiol. 2000 Jan;21(1):124-9.

9. Roehrig J. The manufacturer's perspective. Br J Radiol. 2005;78 Spec No 1:S41-5.

10. Johnston K, Brown J. Two view mammography at incident screens: cost effectiveness analysis of policy options. BMJ. 1999 Oct 23;319(7217):1097-102.

11. Laming D, Warren R. Improving the detection of cancer in the screening of mammograms. J Med Screen. 2000;7(1):24-3

12. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med. 2007 ;356(14):1399-409

13. Wigton R. Social Judgement Theory and Medical Judgement. Thinking and Reasoning 2(2/3) 175-100.

14. Tversky A and Kahneman D Judgment under Uncertainty: Heuristics and Biases Science 1974: 185(4157): 1124 – 1131

15. Harries, C., Evans, J.St.B.T. and Dennis, I. Measuring Doctors' Self-Insight into their Treatment Decisions. Applied Cognitive Psychology 2000, 14, 455-477.

16. Graber M, Gordon R, Franklin N. Reducing diagnostic errors in medicine: what's the goal? Acad Med. 2002 Oct;77(10):981-92.

17. Swensson RG, Hessel SJ, Herman PG. The value of searching films without specific preconceptions. Invest Radiol. 1985 Jan-Feb;20(1):100-14.

18. Berbaum KS, Franken EA Jr, Dorfman DD, Barloon TJ. Influence of clinical history upon detection of nodules and other lesions. Invest Radiol. 1988 Jan;23(1):48-55.

19. Loy CT, Irwig L. Accuracy of diagnostic tests read with and without clinical information: a systematic review. JAMA. 2004 Oct 6;292(13):1602-9

20. Dawes TJ, Vowler SL, Allen CM, Dixon AK Training improves medical student performance in image interpretation Br J Radiol. 2004;77(921):775-6.

21. Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, Conant EF. How experience and training influence mammography expertise. Acad Radiol. 1999 Oct;6(10):575-85

22. Mello-Thoms C. Perception of breast cancer: eye-position analysis of mammogram interpretation. Acad Radiol. 2003 Jan;10(1):4-12.

23. Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. J Natl Cancer Inst. 2003;95(4):282-90.

24. Barlow WE, Chi C, Carney PA, Taplin SH, D'Orsi C, Cutter G, Hendrick RE, Elmore JG. Accuracy of screening mammography interpretation by characteristics of radiologists. J Natl Cancer Inst. 2004;96(24):1840-50.

25. Smith-Bindman R, Chu P, Miglioretti DL, Quale C, Rosenberg RD, Cutter G, Geller B, Bacchetti P, Sickles EA, Kerlikowske K. Physician predictors of mammographic accuracy. J Natl Cancer Inst. 2005;97(5):358-67.

26. Esserman L, Cowley H, Eberle C et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst. 2002;94(5):369-75.

27. Wagner RF, Beam CA, Beiden SV. Reader variability in mammography and its implications for expected utility over the population of readers and cases. Med Decis Making. 2004;24(6):561-72.

28. Bartlett, F.C. Remembering: An Experimental and Social Study. Cambridge University Press, Cambridge, UK.1932

29. Schank R. and Abelson R. Scripts, Plans, Goals and Understanding, Erlbaum, 1977

30. Lesgold A. Rubinson H, Feltovitch P, Glaser R. Klopfer D. Wang Y. Expertise in a complex skill: diagnosing X-ray pictures. In Chi MTH, Glaser R, Farr MJ (ed) The Nature of Expertise. LEA, Hillsdale, New Jersey, USA. 1988 p. 311-42

31. Norman GR. The epistemology of clinical reasoning: perspectives from philosophy, psychology, and neuroscience. Acad Med. 2000 Oct;75(10 Suppl):S127-35

32. Norman G. Research in clinical reasoning: past history and current trends. Med Educ. 2005 Apr;39(4):418-27

33. Kulatunga-Moruzi C, Brooks LR, Norman GR. Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. Teach Learn Med. 2001;13(2):110-6

34. Ark TK, Brooks LR, Eva KW. Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices? Acad Med. 2006;81(4):405-9.

35. Kulatunga-Moruzi C, Brooks LR, Norman GR. Using comprehensive feature lists to bias medical diagnosis. J Exp Psychol Learn Mem Cogn. 2004 May;30(3):563-72.

36. Kundel H. Visual Search in Medical Images in Beutel J, Kundel HL and Van Metter RL (Eds.) *The Handbook of Medical Imaging: Volume 1 Physics and Psychophysics* SPIE Press, Bellingham, Washington USA, 2000 pp. 837-893.

37. Nodine CF and Mello-Thoms C. The Nature of Expertise in Radiology in Beutel J, Kundel HL and Van Metter RL (Eds.) *The Handbook of Medical Imaging: Volume 1 Physics and Psychophysics* SPIE Press, Bellingham, Washington USA, 2000 pp. 859-894.

38. Kundel HL, Nodine CF, Carmody DP. Visual scanning, pattern recognition and decision making in pulmonary nodule detection. *Invest Radiol* **1978**;13:175 -181

39. Kundel H and La Follette PS. Visual Search Patterns and Experience with Radiological Images. Radiology. 1972 103: 523-528.

40. Mello-Thoms C, Hardesty L, Sumkin J, Ganott M, Hakim C, Britton C, Stalder J, Maitz G. Effects of lesion conspicuity on visual search in mammogram reading. Acad Radiol. 2005 Jul;12(7):830-40.

41. Kundel H. and Nodine CF. Interpreting chest radiographs without visual search. Radiology. 1975. 116:527-532.

42. Mugglestone M, Gale AG, Cowley HC, Wilson ARM. Diagnostic performance on briefly presented mammographic images. Proceedings SPIE Medical Imaging. 1995 v 2436: 106-15.
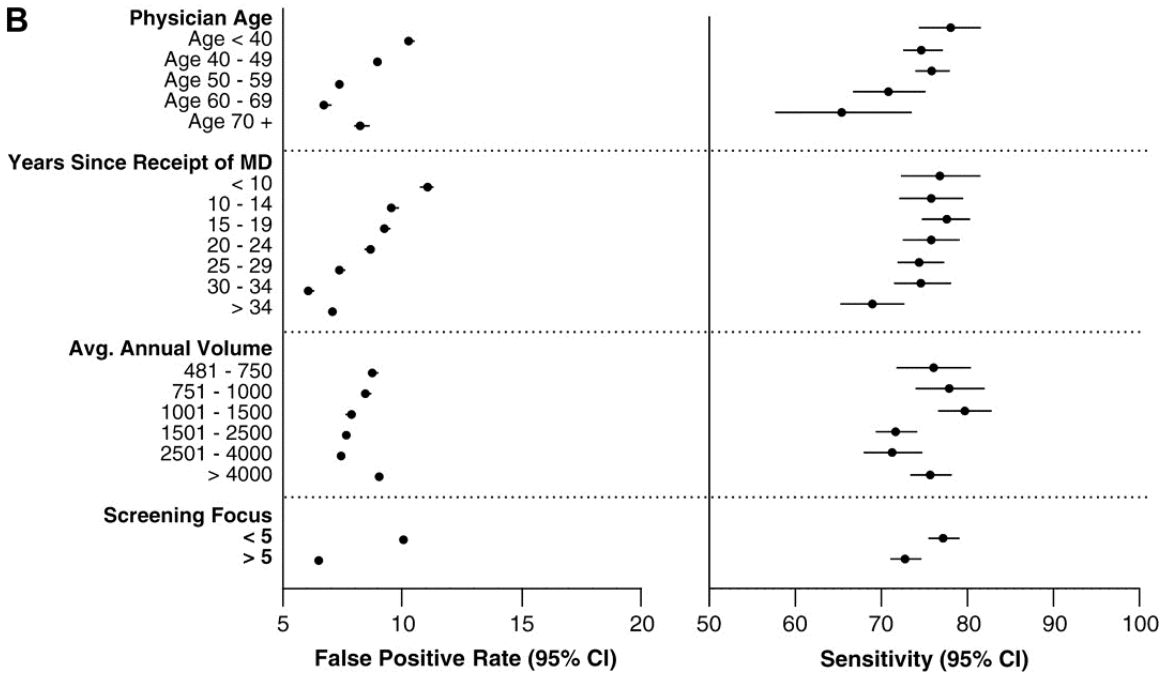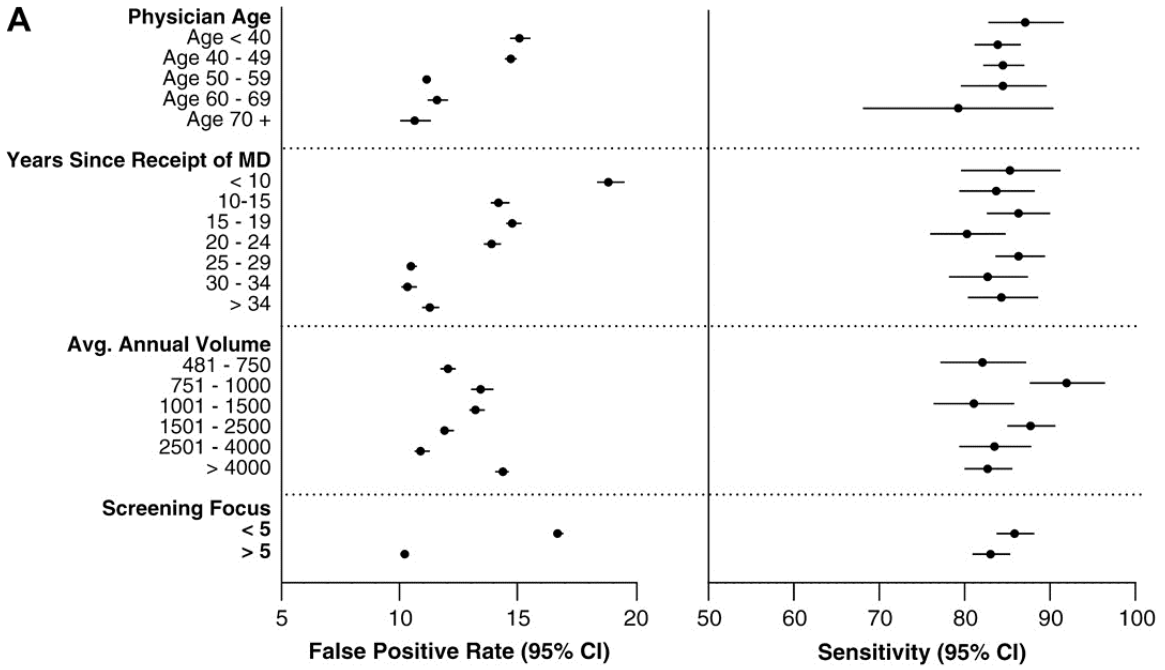
43. Christensen EE, Murry RC, Holland K, Reynolds J, Landay MJ, Moore JG. The effect of search time on perception. Radiology. 1981 Feb;138(2):361-5.

44. Nodine CF, Mello-Thoms C, Kundel HL, et al. Time course of perception and decision making during mammographic interpretation. AJR Am J Roentgenol. 2002 Oct;179(4):917-23.

45. Mello-Thoms C, Dunn SM, Nodine CF, Kundel HL. The perception of breast cancers: a spatial frequency analysis of what differentiates missed from reported cancers. IEEE Trans Med Imaging. 2003;22(10):1297-306

46. Marr, D. Vision. W. H. Freeman. 1983

47. Ahissar M, Hochstein S. The reverse hierarchy theory of visual perceptual learning. Trends Cogn Sci. 2004 Oct;8(10):457-64

48. Sowden PT, Davies IR, Roling P. Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity? J Exp Psychol Hum Percept Perform. 2000 Feb;26(1):379-90

49. Haller S, Radue EW. What is different about a radiologist's brain? Radiology. 2005 Sep;236(3):983-9.

50. Bar M, Kassam KS, Ghuman AS et al. Top-down facilitation of visual recognition. Proc Natl Acad Sci U S A. 2006;103(2):449-54.

51. Thompson KG, Biscoe KL, Sato TR (2005) Neuronal basis of covert spatial attention in the frontal eye field. J Neurosci 25:9479–9487

52. Treue S. Visual attention: the where, what, how and why of saliency. Curr Opin Neurobiol. 2003 Aug;13(4):428-32.

53. Thompson KG, Bichot NP, Sato TR. Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. J Neurophysiol. 2005 Jan;93(1):337-51

54. Collins J. Medical education research: challenges and opportunities. Radiology. 2006 Sep;240(3):639-47.

55. Sharples M, Jeffery NP, du Boulay B, Teather BA, Teather D, du Boulay GH. Structured computer-based training in the interpretation of neuroradiological images. Int J Med Inform. 2000;60(3):263-80.
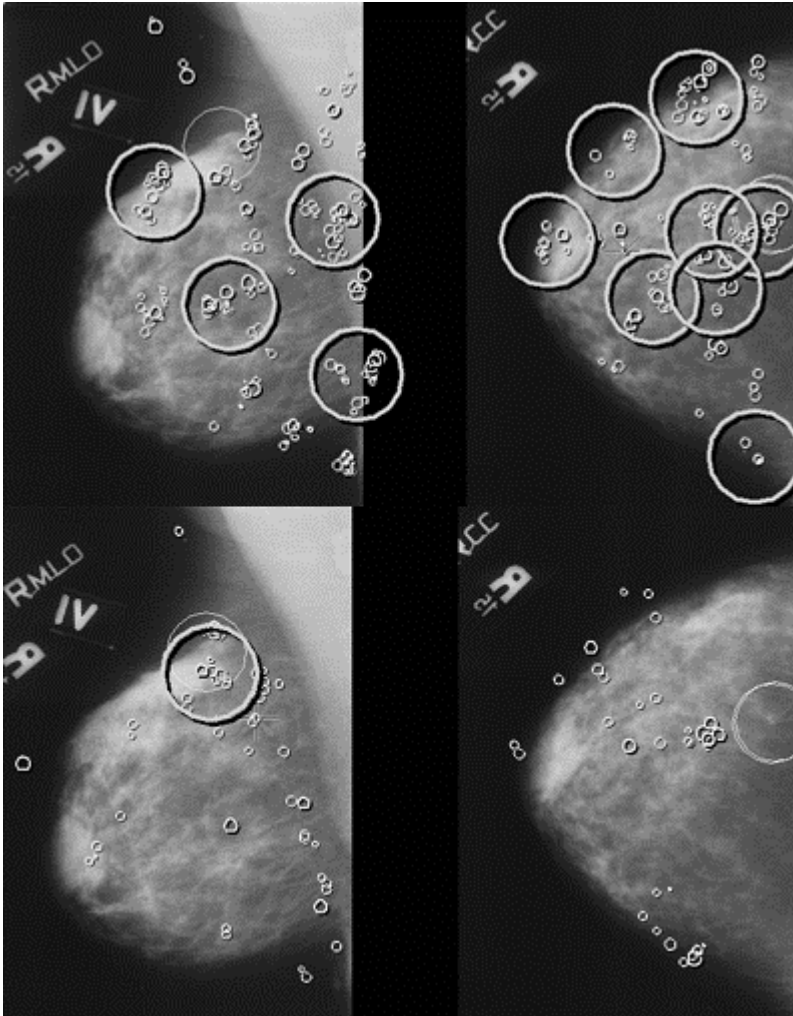
56. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 2004 Oct;79(10 Suppl):S70-81.

57. Lloyd S, Jirotka M, Simpson AC, Highnam RP, Gavaghan DJ, Watson D, Brady JM. Digital mammography: a world without film? Methods Inf Med. 2005;44(2):168-71

58. Gentili A, Chung CB, Hughes T. Informatics in radiology: use of the MIRC DICOM service for clinical trials to automatically create teaching file cases from PACS. Radiographics. 2007 ;27(1):269-75.

59. Kundel HL, Nodine CF, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: gaze-tracking study. Radiology. 2007 Feb;242(2):396-402.
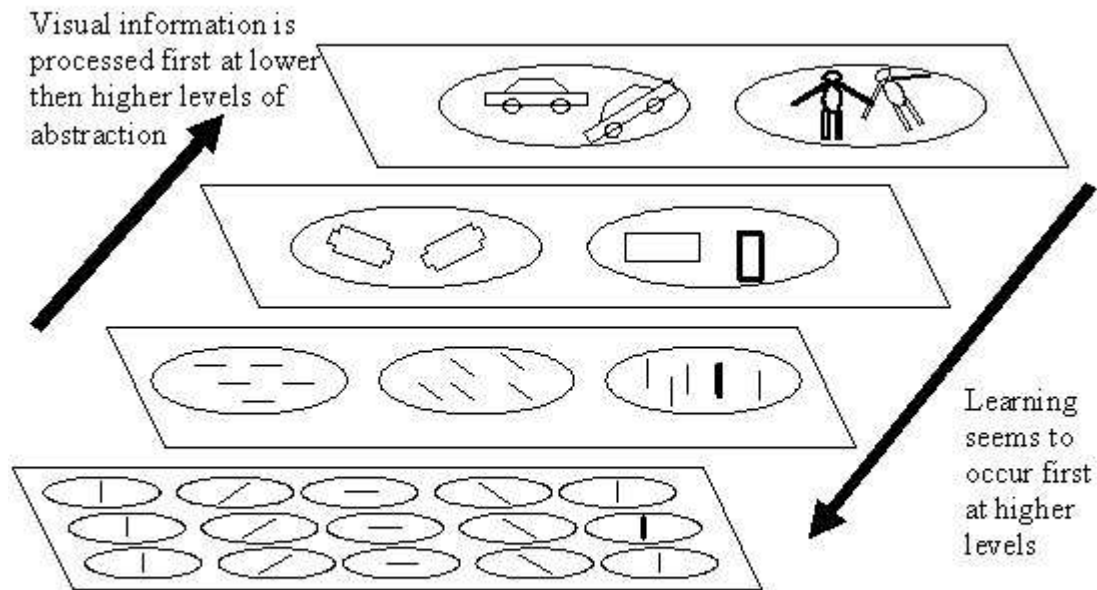
Figure/Table Captions

Caption for Figure 1. 95% confidence intervals for false positive rate and sensitivity of radiologists classified by age, years since qualification (greater or less than 5 years), annual volume and whether or not they focussed on screening, for first (A) and subsequent screening visits (Reprinted by permission of Oxford University Press).

A

Physician Age
Age < 40
Age 40 - 49
Age 50 - 59
Age 60 - 69
Age 70 +

Years Since Receipt of MD
< 10
10-15
15 - 19
20 - 24
25 - 29
30 - 34
> 34

Avg. Annual Volume
481 - 750
751 - 1000
1001 - 1500
1501 - 2500
2501 - 4000
> 4000

Screening Focus
< 5
> 5

False Positive Rate (95% CI)

Sensitivity (95% CI)

B

Physician Age
Age < 40
Age 40 - 49
Age 50 - 59
Age 60 - 69
Age 70 +

Years Since Receipt of MD
< 10
10 - 14
15 - 19
20 - 24
25 - 29
30 - 34
> 34

Avg. Annual Volume
481 - 750
751 - 1000
1001 - 1500
1501 - 2500
2501 - 4000
> 4000

Screening Focus
< 5
> 5

False Positive Rate (95% CI)

Sensitivity (95% CI)

Caption for Figure 2. Example of the scan path of two radiologists reading the same case. The small circles show the location of the fixations, the light circles show the location of a malignant lesion and the dark circles the areas that attracted more than 1s of visual dwell. (© 2003 IEEE)

Caption for Figure 3: The Reverse Hierarchy Theory of Perceptual Learning. The processing of visual information moves through a hierarchy of levels, with lower levels being responsible for detecting simple image features such as lines or edges, and higher levels performing the recognition or classification of objects. If perceptual learning resulted from changes in the processing of low-level features we would expect the learning to be specific to the low-level features of the stimulus (such as contrast or orientation). Such low-level learning seems to occur, but only after higher level learning has taken place and only if task demands are high (for example if the images are noisy).

Caption for Figure 4:

This diagram shows how innovative presentation can help teach classification. The tool displays a two dimensional projections of a conceptual space, created by charaterising cases along different axes. This can help trainees to get a sense of how instances of a disease are distributed around a typical case. The distance between a case of unknown diagnosis and the typical presentations of candidate diagnoses can be used to help in classification.