Reader error, object recognition, and visual search

Harold L. Kundel

Department of Radiology, University of Pennsylvania School of Medicine
3600 Market St. Suite 370 Philadelphia, PA 19104

ABSTRACT

Small abnormalities such as hairline fractures, lung nodules and breast tumors are missed by competent radiologists with sufficient frequency to make them a matter of concern to the medical community; not only because they lead to litigation but also because they delay patient care. It is very easy to attribute misses to incompetence or inattention. To do so may be placing an unjustified stigma on the radiologists involved and may allow other radiologists to continue a false optimism that it can never happen to them. This review presents some of the fundamentals of visual system function that are relevant to understanding the search for and the recognition of small targets embedded in complicated but meaningful backgrounds like chests and mammograms. It presents a model for visual search that postulates a pre-attentive global analysis of the retinal image followed by foveal checking fixations and eventually discovery scanning. The model will be used to differentiate errors of search, recognition and decision making. The implications for computer aided diagnosis and for functional workstation design are discussed.

Keywords: Human error, visual search, visual perception, eye movement, mammography, breast cancer, chest radiography, lung cancer

1. READER ERROR

1.1 Error and variation in diagnostic imaging

Periodically, radiologists[1,2], other medical practitioners[3], lawyers[4], and the general public discover that imaging diagnosis is fallible. It might be comforting to know that radiologists are not alone. Most medical tests that have been studied systematically have an error rate[5,6]. In fact almost every activity in which humans make observations is prone to error. Stigler[7] traces the formal study of observational error and variation to 18th century astronomers who set navigational clocks by observing when certain stars crossed the meridian. Smith[2] who developed an anecdotal classification of radiological error, traces attempts to classify human error to Browne and Bacon in the 17th century. Formal interest in understanding and quantifying error in radiology began in the late 1940s with studies of screening for tuberculosis[8,9] and eventually led to the introduction of the receiver operating characteristic (ROC) analysis as a method for assessing and understanding reader performance[10,11]. Screening trials have continued to supply good material for studies of observer error because 1) they focus on a single disease like cancer and even in large population trials investigators attempt to arrive at ground truth. The results of screening trials can be used to estimate the magnitude of error and the degree of variation. The deeper causes for error are not well understood and analyses of the errors made by radiologists is usually purely anecdotal [12,13]. The analysis that follows is based on studies of projection radiography for breast and lung cancer screening. Breast cancer screening still uses projection radiographs. Because past trials of lung cancer screening using projection chest radiography have had controversial results[14], computed tomography is now being advocated[15].

1.2 Quantifying error in lung and breast cancer screening by projection radiography

The sensitivity and specificity of readings of mammograms made between 1980 and 1987 in the Canadian National Breast Screening Study were reported as 75 and 94 percent respectively[16]. In a follow-up study, an expert “reference radiologist” reviewed 5200 mammograms that were selected randomly from among the screening examinations of women who were known not to have cancer and 575 screening mammograms on which cancer was detected[17]. The joint agreement of the original readers and the reference reader with truth data were used to construct the curves in Figure 1.

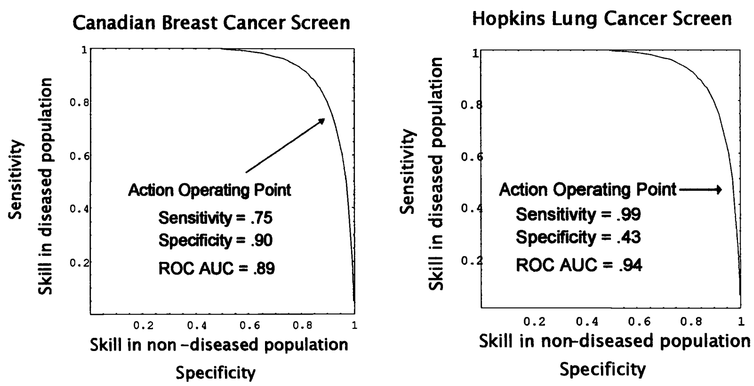
Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment,
edited by Dev P. Chakraborty, Miguel P. Eckstein, Proceedings of SPIE Vol. 5372
(SPIE, Bellingham, WA, 2004) · 1605-7422/04/$15 · doi: 10.1117/12.542717

The curves are plots of diagnostic skill in diseased populations (sensitivity) against diagnostic skill in non-diseased populations (specificity) and are called "diagnostician operating choices" (DOC) curves by C. Beam[18]. The expected operating point was determined from the reported sensitivity and specificity[16].

Figure 1 also shows the DOC curves estimated from the pooled data from a lung cancer screening program sponsored by the National Cancer Institute. About 30,000 male volunteers over age 45 who smoked at least one pack of cigarettes a day were enrolled between 1971 and 1978 at three institutions - Johns Hopkins, Memorial Sloan-Kettering, and the Mayo Clinic. Each candidate filled out a questionnaire, received chest radiography and gave a pooled sputum sample for cytology. The volunteers were followed for five years so that most of the cases of lung cancer present at the inception of the study were eventually identified. The action data on the initial screening and five years of follow-up were used to estimate DOC curves[19-22].

Figure 1. DOC curves for breast and lung cancer screening by projection images

The difference in the action operating point for breast and lung screening probably occurred because mammographers are more tolerant of callbacks.



- Sensitivity and specificity, which are frequently (naively) combined to express an error rate, will vary with the operating point on the DOC curve.
- Although not explicitly shown here because the data from individuals are pooled, different readers will have different operating points producing wide statistical variability[23].
- The area under the ROC curve (AUC) is not sufficient to fully understand performance. Even though AUC is greater for chest than breast screening (.94 to .89), the sensitivity at the action based operating point is different (.43 to .75).
- In screening for a disease with low prevalence the specificity is very important in setting the operating point because it determines the number of callbacks that require intervention.

1.3 The Taxonomy of Error

Historically, radiologists have been concerned with understanding error, more as a way of improving clinical practice than as an avenue for understanding themselves. A classification of error is shown in Table 1. It follows a serial model of what Russell Morgan[24] called a "visual imaging system", that is an imaging system in which the human observer views a displayed version of the image and provides the readout.

Table 1. Classification of error for the detection of discrete lesions embedded in structured backgrounds

Type of Error	Explanation
Technological	Inadequate Lesion Conspicuity
Perceptual	
Visual Search	Inadequate scanning strategy or low peripheral conspicuity of lesion
Recognition	Lack of knowledge and experience with lesion and background properties
Satisfaction of Search	Possible failure of vigilance mechanism
Covert Decision	Detect but do not perceive lesion
Cognitive	
Classification	Detect, perceive and misclassify lesion

Error is assigned to one of three stages: technological, perceptual and cognitive. Technological error will not be discussed here. The model for the perceptual and cognitive stages is drawn largely from the work of perceptual psychologists as summarized by Gregory[25] and Rock[26]. Perceptual error is divided into visual search, object recognition and covert decision while cognitive error deals with the classification of objects that are recognized. Eye tracking studies during the search for discrete lesions in projection chest images and mammograms have contributed to the understanding of perceptual and cognitive errors.

2. OBJECT RECOGNITION

The brain is one of the most complicated structures in the universe and any model of its function is bound to be utterly simplistic. Nevertheless, perceptual psychologists seem to enjoy developing models that can be described with flowcharts. The error classification described in table 1 is derived from a model that has three major components.

- Development of a literal perception
- Construction of a preferred perception
- Analysis of the preferred perception

The *literal perception* is developed automatically without any dependence on prior experience or expectation in the initial few tens of milliseconds of viewing. It is developed from visual primitives obtained from the retinal neural network. The visual primitives contain information about boundaries from luminance, color, texture, motion and disparity[27]. The perceptual system then constructs a *preferred perception* using a statistical knowledge base that assigns object descriptions to all of the structures in the image[28]. The preferred perception can be developed in a few hundred ms or it can develop more slowly requiring a few seconds. It can also be modified. Ambiguities detected in the peripheral visual field during the initial few ms of viewing may be examined in detail by shifting the gaze axis. In addition, the perceptual system may deliberately search the image for objects that are not peripherally conspicuous and consequently not immediately identifiably by the peripheral vision. If they are found, they are added to the preferred perception.

The preferred perception is used for action. In the everyday world, it might keep the observer from bumping into a lamppost while walking. In diagnostic imaging the objects identified in the preferred perception are deconstructed and classified. The deconstruction consists of describing features that are thought to be inherent in the image but that may actually be perceptual constructions. Classification of a tumor as benign or malignant may be justified by describing the features of the object. In this model, features are not extracted until after the preferred perception has formed.

- The extraction of features may change the preferred perception but they are essentially a cognitive construct - made up to explain the appearance in the preferred perception.

3. EYE MOVEMENT AND VISUAL PERCEPTION

3.1 Acquisition of the image by the retina

The retina is a complex, highly organized but non-uniform image receptor array. The individual sensory elements, called rods and cones because of their microscopic appearance, have different sensitivity to the spectrum and the intensity of light and are non-uniformly distributed over the surface of the retina. The rods are very sensitive to light requiring only a few quanta for excitation while the cones require many quanta. The rods are monochromatic and most sensitive to short wavelength (blue) light but the cones come in three varieties that are sensitive to short (blue), medium (green) and long (red) wavelengths of light. Together they produce color vision. The cones are most closely packed in a small, saucer like depression in the retina called the fovea centralis, which is located on the optical axis of the eye. Only cones can be found in the very center of the fovea. The density of cones then decreases as the periphery is approached.
Figure 2 is a simplified diagram of the retinal neural network, which consists of 3 layers of cells with two intermediate layers of synapses. The signal from multiple sensory cells is relayed to the bipolar cells and then to the ganglion cells whose axons form the optic nerve. Amacrine cells in the middle layer interconnect the sensory and ganglion cells. Notice that multiple rods and cones provide the signal for a single ganglion cell.

In the fovea a single cone may be associated with a single bipolar and ganglion cell but in the periphery the signal from multiple cones stimulates one ganglion cell. One consequence of the interconnections is that the retinal sensory array can be modeled as a series of overlapping sensory fields with specific functions. Another consequence is that visual acuity is greatest at the center of the visual field and decreases steeply toward the periphery.

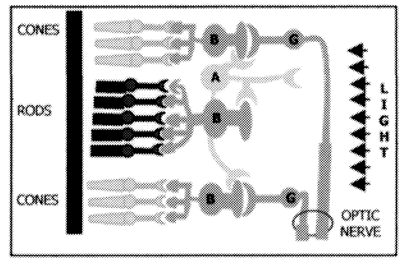


Figure 2. A simplified and idealized diagram of the retina

Three cell layers contain the sensory cells (rods and cones), the bodies of the bipolar cells (B) and the cross-linking amacrine cells (A), and the bodies of the ganglion cells (G), the axons of which make up the optic nerve. There are two intermediate layers of synapses. Note that the light enters from ganglion cell side. The vertical black line represents the outer pigmented layer.

The relative visual acuity as determined by behavioral measurements is less than 50 percent by 1 degree of visual angle and less than 5 percent by 10 degrees. Figure 3 shows the relative visual acuity as a function of distance from the center of the fovea and on a chest image viewed at 57 cm (roughly arm's length), the size of the 1 degree rod free zone, the 5 degree fovea and the 10 degree field of view.

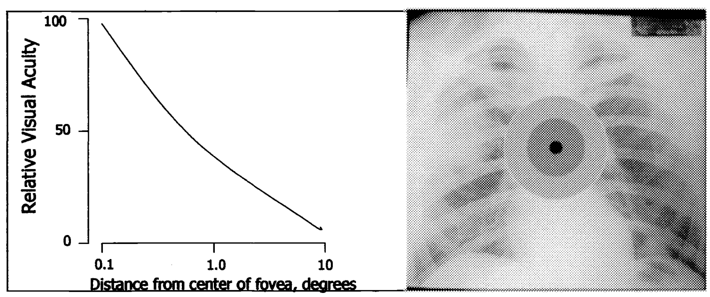


Figure 3. The relative visual acuity and the relative size of the associated field of view

Note that the distance is on a logarithmic scale.

The three concentric disks on the chest image show the size of a 1, 5 and 10 degree visual field when viewed at 57 cm distance, which is roughly arm's length.

3.2 Some basics of eye movement

The human field of view is very large extending more than 180 degrees. Although the entire field of view is perceived as equally sharp, details can only be seen in the very center of the field. This can be easily demonstrated by looking at a word in the center of a printed page and trying to read the surrounding words without moving the eyes, the head or the page. Normally people first move their eyes and then their heads to bring objects of interest into the center of the visual field. The eyes move from place to place in rapid jumps (saccades). The intervening fixations have a median duration of about 300 ms. The eyes are never entirely stationary. During steady fixation there are fine oscillations called microsaccades and a slow drift of the optical axis away from the initial landing point. Constant movement is a physiological necessity. The retina responds to changes in illumination and an image that is stabilized on the retina fades away. An eye that does not move cannot see. Prolonged fixation is rare. When the gaze is directed steadily at one place, closely spaced fixation clusters of 200 to 300 ms instead of one long fixation are observed.

- Investigators who use eye position recording as a method for studying perception generally believe that eye position is entrained by the viewers' attention.
- The viewer is generally unaware of either the position of the axis of the gaze or the track or scanpath that the gaze followed during viewing.

3.3 Visual scanpaths

The movement of the eyes can be tracked by a number of devices, some more intrusive than others. Very clever analog devices that record a light beam reflected from a tiny mirror attached to the cornea using a suction cup have produced very accurate recordings. Of course the head must be immobilized[29]. Modern devices use infra-red light reflected from the cornea and location of the center of the pupil in a television image to track the eye and a magnetic recording device to track the head. With proper calibration, accuracy of less than one degree of visual angle can be obtained with minimal disturbance of the observer. Position data are sampled about 60 times a second.

Sequential fixation location and duration can be computed from the sampled data. The scanpath is a plot of the sequence of fixations on the image. Examples of a scanpaths are shown in figure 4. The exact scanpath is determined by the structure of the image and by the experience and expectancies of the viewer.

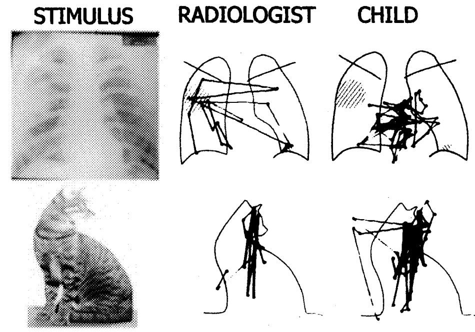


Figure 4. Visual scanpaths on an image of a chest and a cat.

A radiologist and a child aged 6 were shown a chest image with two infiltrates in the lungs and a picture of a cat and instructed tell us what they saw. The radiologist reported a cat and a chest with two sites of pneumonia. The child reported a cat and "somebody's bones". The scanpaths on the cat are very similar and on the chest are very different.

Yarbus[29] very elegantly demonstrated that the task given the viewer modifies the scanpath. This has also been shown for radiologists viewing chest images[30]. The scanpaths of the child and the radiologist shown in figure 4 are an extreme example of the effect of knowledge and expectancy. Scanpaths change during medical education[31]. The most noticeable change occurs early in medical school after the medical students have learned about anatomy and gross pathology as shown in Figure 5.

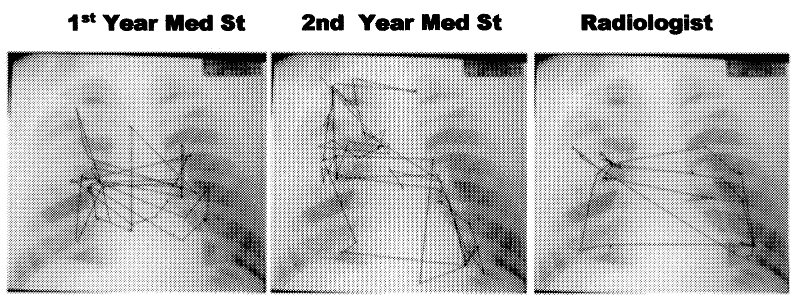


Figure 5. The scanpath of a first year medical student, a second year medical student and a radiologist looking at an image of a projection chest with two infiltrates.

Visually, the scanpath of the second year medical student is more like the radiologist than that of the first year medical student.

- Understanding pathology is very important for recognizing diagnostically important objects in the perceived medical image.

3.4 Fixation distribution and cluster dwell time

The scanpath shows the sequence in which information about details in the image was collected. The same details can be accessed using different scanpaths as shown in figure 6. Visually scanpaths on chests[30], on the bones of the extremities[32] and on mammograms can be roughly classified as circumferential, zigzag and complex. In addition, visual fixations tend to cluster on informative details in the image[33,34]. Inspection of the scanpaths of the 2nd year medical student and the radiologist in figure 5 shows a concentration of fixations on the infiltrates. Although scanpaths are interesting, it has been difficult to analyze them objectively and we have concentrated on the analysis of fixation cluster dwell time.

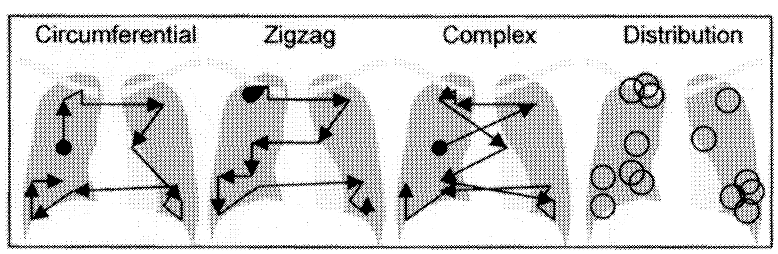


Figure 6 Three scanpath patterns that produce the same fixation distribution

4. VISUAL SEARCH

4.1 Searching for discrete targets in medical images

Visual search is a common human activity. We search for a book on a bookshelf, a car in the parking lot, or a face in a crowd. Each of the many possible individual search tasks has some elements in common.

- A search object, called the target, with properties that are either known exactly or statistically
- Objects that can be confused with the search object (mimicry)
- A structured background of varying complexity (camouflage)
- A well defined search field that is generally large compared with the size of the target

Targets in medical images are embedded in anatomical backgrounds. In projection chest images small tumors in the lung are embedded in the shadows of ribs and blood vessels. Only the statistical properties of the tumors are known and the chest image backgrounds are variable even within the same individual. Technical fact such as positioning, respiration and projection angles can materially change the appearance of the image. Masses and microcalcifications are similarly camouflaged on mammograms. Hairline fractures in the bones on projection images of the extremities are very fine details and may not perturb the literal perception sufficiently to attract foveal attention.

The studies of visual search conducted in my laboratory developed from a larger effort aimed at using image processing to improve the visibility of lung tumors and thereby lower error rates[35]. The basic approach was to have a group of radiologists report on a set of processed and unprocessed chest images half of which contained an inconspicuous lung nodule. The detection performance with and without image processing was then compared to determine if the image processing did any good. It was devilishly difficult to find inconspicuous lung nodules by searching through clinical images and, instead, nodules were synthesized photographically on otherwise normal chest images[35]. This methodology, which gave us exquisite control over target properties and target location, not only simplified the study of image processing but also provided material for the study of the sources of error. One of the sources was search; either incomplete search, which implies a defective search strategy, or the more insidious "satisfaction of search" which implies a discontinuation of search when the observer is satisfied about the meaning of the image[36]. Both of these mechanisms were highly speculative and we decided to study them more rigorously, in other words, to collect data.

Search can be studied in two ways: 1) using the behavioral measures of reaction time and accuracy and 2) recording the location of the gaze axis. Reaction time is a measure of how long it takes to report a target. Accuracy can be measured as the proportion of targets detected or, preferably, as an index of detectability that considers the impact of both true positive and false positive reports. The location of the gaze axis can be measured by recording eye and head motion using a system that is calibrated to project the locus of the axis into the display plane. In my opinion, there is an uncertainty principle asserting that reaction time and gaze location at the instant of detection cannot be determined in the same experiment. If asked to push a button as soon as a target is detected, a reader will almost always direct the gaze to the location that is being reported and then push the button. So the gaze is always directed at the detection site at the time of the button press even though the detection may have occurred before the site was fixated directly. In visual search experiments as opposed to card tricks "the eye is quicker than the hand".

My colleague Cal Nodine and I believe that search has four main components[37].

- Global Impression - Examination of the total visual scene for image perturbations
- Foveal Verification - Direct examination of perturbations by the foveal vision
- Discovery Scanning - Deliberate scanning of the foveal vision over the search field
- Reflective Search - Iterative scanning to affirm and support initial decisions

4.2 Searching for lung nodules in projection chest images: Scanning, recognition and decision errors.

Our initial studies of readers searching chest images for one cm diameter, simulated lung nodules showed that most of them were fixated by the foveal vision *even though the viewers denied seeing them.*[38] The analysis was conducted by determining the location of the axis of the gaze relative to the center of the nodule.

The five degree circular region around the locus of the gaze axis was designated as the *useful visual field.* Any fixation within 2.8 degrees of the nodule center was scored as a hit. This value was chosen because a study of 117 true positive trials showed that 90% of the nodules were hit by a 2.8 degree radius field. The dwell time of the useful visual field on the nodule was also calculated. If the nodule was not hit, a scanning error was scored. If the nodule was hit for one fixation and the gaze did not return, a recognition error was scored. If the gaze was prolonged or if the gaze returned to the target one or more times a decision error was scored. The situation is illustrated in figure 7. Analysis of 20 false negative trials showed 30 percent scanning errors, 25 percent recognition errors and 45 percent decision errors[38]. The proportion of the errors in the scanning and recognition class can be changed by changing the size of the useful visual field but the basic classification is unchanged.

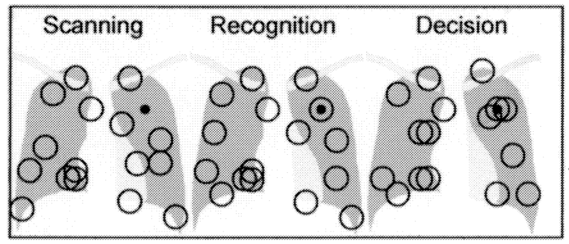


Figure 7. Three classes of error as determined by analysis of the scanning of a 5 degree diameter field over the image.

The black spot in the left upper lung region represents the nodule. Each circle represents one fixation. Notice that some of the circles tend to form clusters.

4.3 Searching for small lesions in projection images: Analysis of visual dwell times.

The ability to detect small lesions with the peripheral vision depends upon a number of factors of which the most important are the size-contrast combination, the border sharpness and continuity, and the camouflage provided by the background locally by overlapping structures and countershading and distantly by mimicry[39,40]. Together these factors make up what has been called target or lesion conspicuity[41-43]. A very conspicuous lesion like a dense calcification (moderate size, high contrast, sharp and well defined boundaries) can be seen well out in the peripheral field and may be perceived, classified and not tagged for a checking fixation; whereas an inconspicuous lesion like a nodule (moderate size, low contrast, blunted and poorly defined boundaries) may be peripherally identified but may require checking fixations for classification; and a very inconspicuous lesion like microcalcifications (small size) might require discovery scanning. I believe that the fixation cluster dwell time depends on both the lesion conspicuity and on the difficulty of the decision either to perceive (covert) or to assign to a class (overt). Analysis of both lung nodule detection and cancer detection on mammograms has been done by comparing cluster dwell time on lesions and on false positive location to the dwell time of the clusters that occur on lesion-free images. Detected lesions and false positive locations receive prolonged fixations and 50 to 60 percent of missed lesions receive prolonged fixation as well[44,45]. An example of the results of a typical study is shown in Table 2.

Table 2. Mean total fixation dwell time on nodules and on corresponding nodule free areas of projection chest images.

The data were extracted from a study of two radiologists searching 120 chest radiographs, half of which contained a solitary lung nodule[44].

Decision Outcome	Number	Mean fixation time, sec	SE
True Positive	49	2.76	.23
False positive	22	2.53	.22
False negative	46	2.44	.15
Suspicious negative	122	2.06	.17
True negative	4878	0.51	.10

4.4 Satisfaction of Search

Satisfaction of search (SOS) is the catchy phrase applied to the phenomenon in which the detection of one abnormality interferes with the detection of another. Tuddenham[46] probably was the first radiologist to suggest that the mechanism consisted of stopping the search for abnormalities when the radiologist was satisfied about the meaning of the image. The existence of SOS has been elegantly demonstrated experimentally by Berbaum and his colleagues[47-50]. The mechanism of SOS has been elusive. Observers do not stop viewing when one abnormality has been found on an image with multiple abnormalities. Eye position studies have shown very little difference between trials in which target lesions were missed with and without the presence of distracter lesions[51]. An increase in the proportion of recognition errors was observed suggesting that the SOS mechanism may have more to do with recognition than scanning.

5. IMPROVING READER PERFORMANCE

5.1 The technological fix

Our original idea about improving observer performance was to show readers where they did not look in an image and to ask them to examine those areas again. The finding that 75 of the missed nodules were actually scanned by a 5 degree visual field and 45 percent received prolonged visual dwell times suggested that asking readers to reevaluate areas where they looked for a prolonged time would be more fruitful. Just as in computer aided diagnosis (CAD) the areas selected for reevaluation included both missed lesions and false positives. A study of visual feedback after search showed a small but statistically significant improvement in detection performance as measured by the area under the alternative free response receiver operating characteristic (AFROC) curve[52,53]. Studies of mammography showed that the gaze duration could be used to identify missed lesions[45] but a study of visual feedback did not show any improvement in performance[54].

The difference between the chest studies and the mammogram studies may have been in the selection of lesions and readers. The chest studies used simulated lesions and residents as observers. The mammogram studies used real lesions and a mixture of experts, fellows (pre-experts) and residents. Comparing the performance of the experts and less-than-experts started us on the road of examining the properties of expertise and of studying how expertise in a visual task is developed.

5.3 Identifying and understanding expertise

It is very hard to objectively define an expert in radiology. Some radiologists have a very high level of pattern recognition skill and remember numerous previous cases and outcomes. Others have well honed clinical deductive skills that are used to arrive at a diagnosis for a particular image. Experts can generally be recognized by their peers although the peers may not be able to stipulate the basis for such recognition.

Experts do score better on objective tests of detection performance and test performance is roughly related to experience[55]. Experts also tend to fixate lesions earlier in search and experts are better able to reject image perturbations that generate false positive responses in non-experts[56,57]

5.4 The futility of teaching trainees how to scan images

Some teachers of radiology advocate a systematic viewing of the images. Garland advocated a three step procedure for the chest, in which the interpreter "inspected first the central cardiovascular shadow and then the lungs, right and left, as a whole. Then he would require his eyes to follow each lung field, space by space, from apex to base; then rib by rib from base to apex. Next he would 'look across' from space to space, comparing right and left sides, from apex to base. Finally, he would review the bony thorax and extrathoracic soft tissues for any apparent abnormality". None of the radiologists that we have studied ever did this. Most looked immediately at the abnormalities (see figures 4 and 5) and the more experienced radiologists typically made a circumferential scan of the image. Carmody et al.[58] found that although radiologists had been taught to be "systematic and directive with comparisons of bilateral features", most employed a free search method. When the search object is known to be very inconspicuous such as a rib fracture or breast microcalcifications, systematic scanning is helpful and perhaps mandatory but otherwise it may be unwise to concentrate on a search strategy. Understanding pathology and the range of normal variation is much more important.

5.5 Understanding normal variation

One of the major difficulties that observers encounter in lung and breast cancer screening is separating true lesions from image perturbations that result in false positive responses[44,59]. It is not clear if the displayed image physical properties of false positive perturbations are intrinsically different from the properties of lesions. Even mammographers have difficulty distinguishing true lesions from false lesions when the true lesions are in the very early stages of developmenty[59]. Additionally, it is not clear exactly what displayed physical properties are important. We have been exploring various ways of characterizing localized regions in mammograms that receive prolonged visual attention[60]. It would be of interest to know what image attributes attract visual attention. If there are definite differences between true lesions and misleading appearances, they might be selectively enhanced to improve observer performance.

5.6 Presenting images to observers

The initial global impression of the image is very important for priming perceptual search and the basic search is completed in a few seconds. This implies that the initial appearance of the displayed image is important for perception and although there is no hard evidence, a confusing initial display that has to be manipulated to achieve a display pleasing to the viewer may be detrimental to performance. It is certainly annoying. There is a need to go beyond the optimization of contrast and detail and to consider the ergonomics of the entire workstation[61] perhaps tailoring the display and the controls of the display to individuals and even to the image being displayed.

6. REFERENCES

1. Garland LH: "Studies on the accuracy of diagnostic procedures". *AJR American Journal of Roentgenology.* **82**: 25-38, 1959.
2. Smith MJ: *Error and variation in diagnostic radiology.* Thomas,Springfield, Ill, 1967.
3. Elmore JG, Wells CK, Lee CH *et al.*: "Variability in radiologists' interpretation of mammograms". *New England Journal of Medicine.* **331**: 1493-1499, 1994.
4. Berlin L and Berlin JW: "Malpractice and radiologists in Cook County, IL: Trends in 20 years of litigation". *American Journal of Roentgenology.* **165**: 781-788, 1995.
5. Koran LM: "The reliability of clinical methods, data and judgments (first of two parts)". *New England Journal of Medicine.* **293**: 642-6, 1975.
6. Koran LM: "The reliability of clinical methods, data and judgments (second of two parts)". *New England Journal of Medicine.* **293**: 695-701, 1975.
7. Stigler SM: *The History of Statistics.* Harvard Univ Press,Cambridge, MA, 1968.
8. Birkelo CC, Chamberlain WE, Phelps PS *et al.*: "Tuberculosis case finding. A comparison of the effectiveness of various roentgenographic and photofluorographic methods". *Journal of the American Medical Association.* **133**: 359-366, 1947.
9. Yerushalmy J: The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows, in Tuddenham WJ: *The Radiologic Clinics of North America.* Philadelphia, W.B.Saunders, 1969, vol. 7, pp 381-392.
10. Lusted LB: "Signal detectability and medical decision making". *Science.* **171**: 1217-1219, 1971.
11. Swets JA, Pickett RM, Whitehead SF *et al.*: "Assessment of diagnostic technologies". *Science.* **205**: 753-759, 1979.
12. Smith MJ: "Errors in diagnostic radiology on the basis of complacency". *AJR American Journal of Roentgenology.* **94**: 689-703, 1965.
13. Renfrew DL, Franken Jr. EA, Berbaum KS *et al.*: "Error in radiology: Classification and lesions in 182 cases presented at a problem case conference". *Radiology.* **183**: 145-150, 1992.
14. Strauss GM: "Measuring effectiveness of lung cancer screening: From consensus to controversy and back". *Chest.* **112**: 216S-228S, 1997.
15. Henschke CI, McCauley DI, Yankelevitz DF *et al.*: "Early lung cancer action project: overall design and findings from baseline screening". *Lancet.* **354**: 99-105, 1999.
16. Baines CJ, McFarlane DV and Miller AB: "Sensitivity and specificity of first screen mammography in 15 NBSS centers". *Can Assoc Radiol J.* **39**: 273-276, 1988.
17. Baines CJ, McFarlane DV and Miller AB: "The role of the reference radiologist. Estimates of inter-observer agreement and potential delay in cancer detection in the national breast cancer screening study". *Investigative Radiology.* **25**: 971-976, 1990.
18. Beam CA: Foundations of a scientific medical interpretation in screening mammography: *Mammography Foundations Conference.* Tampa, FL, Moffitt Cancer Center & Research Institute, 2004.
19. Berlin NI, Buncher CR, Fontana RS *et al.*: "The National Cancer Institute cooperative early lung cancer detection program". *Am Rev Resp Dis.* **130**: 545-549, 1984.
20. Flehinger BJ, Melamed MR, Zaman MB *et al.*: "Early lung cancer detection: Results of the initial (prevalence) radiologic and cytologic screening in the Memorial Sloan-Kettering study". *Am Rev Resp Dis.* **130**: 555-560, 1984.
21. Fontana RS, Sanderson DR, Taylor WF *et al.*: "Early lung cancer detection: Results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic study". *Am Rev Resp Dis.* **130**: 561-565, 1984.
22. Frost JK, Ball WC, Levin ML *et al.*: "Early lung cancer detection: Results of the initial (prevalence) radiologic and cytologic screening in the Johns Hopkins study". *Am Rev Resp Dis.* **130**: 549-554, 1984.
23. Beam CA, Layde PM and Sullivan DC: "Variability in the interpretation of screening mammograms by US radiologists". *Arch Intern Med.* **156**: 209-213, 1996.
24. Morgan RH: "Visual perception in fluoroscopy and radiography. Annual oration in memory of John D. Reeves, Jr., M. D., 1924-1964". *Radiology.* **86**: 403-16, 1966.
25. Gregory RL: *The Intelligent Eye.* McGraw-Hill,New York, 1970.
26. Rock I: *Perception.* W.H. Freeman and Company,New York, NY, 1984.

27. Livingstone M and Hubel D: "Segregation of form, color, movement, and depth: anatomy, physiology, and perception". *Science.* **240**: 740-749, 1988.
28. Simoncelli EP and Olshausen BA: "Natural image statistics and neural representation". *Annu Rev Neurosci.* **24**: 1193-216, 2001.
29. Yarbus AL: *Eye Movements and Vision.*Plenum Press,New York, 1967.
30. Kundel HL and Wright DJ: "The influence of prior knowledge on visual search strategies during the viewing of chest radiographs". *Radiology.* **93**: 315-320, 1969.
31. Kundel HL and LaFollette PS: "Visual search patterns and experience with radiological images". *Radiology.* **103**: 523-528, 1972.
32. Hu CH, Kundel HL, Nodine CF *et al.*: "Searching for bone fractures: A comparison with pulmonary nodule search". *Academic Radiology.* **1**: 25-32, 1994.
33. Mackworth NH and Morandi A: "The gaze selects informative details within pictures". *Perception and Psychophysics.* **2**: 547-548, 1967.
34. Kundel HL: "Visual sampling and estimates of the location of information on chest films". *Investigative Radiology.* **9**: 87-93, 1974.
35. Kundel HL, Revesz G and Stauffer HM: "Evaluation of a television image processing system". *Investigative Radiology.* **3**: 44-50, 1968.
36. Tuddenham WJ: "Roentgen image perception - A personal survey of the problem". *Radiologic Clinics of North America.* **3**: 499-501, 1969.
37. Nodine CF and Kundel HL: The cognitive side of visual search in radiology, in O'Regan JK andLevy-Schoen A: *Eye Movements: From Physiology to Cognition.* North Holland, Elsevier Science Publishers, 1987, pp 573-582.
38. Kundel HL, Nodine CF and Carmody DP: "Visual scanning, pattern recognition, and decision making in pulmonary nodule detection". *Investigative Radiology.* **13**: 175-181, 1978.
39. Behrens RR: "On visual art and camouflage". *Leonardo.* **11**: 203-204, 1978.
40. Kundel HL: "Predictive value and threshold detectibility of lung tumors". *Radiology.* **139**: 25-29, 1981.
41. Engel FL: "Visual conspicuity, directed attention and retinal locus". *Vision Research.* **11**: 563-567, 1971.
42. Kundel HL and Revesz G: "Lesion conspicuity, structured noise, and film reader error". *American Journal of Roentgenology.* **126**: 1233-1238, 1976.
43. Revesz G: "Conspicuity and uncertainty in the radiographic detection of lesions". *Radiology.* **154**: 625-8, 1985.
44. Kundel HL, Nodine CF and Krupinski EA: "Searching for lung nodules: Visual dwell indicates locations of false-positive and false-negative decisions". *Investigative Radiology.* **24**: 472-478, 1989.
45. Krupinski EA and Nodine CF: Gaze duration predicts the locations of missed lesions in mammography, in Gale AG, Astley SM, Dance DR *et al.*: *Digital Mammography.* Amsterdam, Elsevier, 1994, pp 399-403.
46. Tuddenham WJ: "Problems of perception in chest roentgenology: facts and fallacies". *Radiologic Clinics of North America.* **1**: 277-289, 1963.
47. Berbaum KS, Franken EA, Jr., Dorfman DD *et al.*: "Satisfaction of search in diagnostic radiology". *Investigative Radiology.* **25**: 133-40, 1990.
48. Berbaum KS, El-Khoury GY and Franken EA, Jr.: "Missed fractures resulting from satisfaction of search effect". *Emergency Radiology.* **1**: 242-249, 1994.
49. Berbaum KS, Franken EA, Jr., Dorfman DD *et al.*: "Cause of satisfaction of search effects in contrast studies of the abdomen". *Academic Radiology.* **3**: 815-26, 1996.
50. Berbaum KS, Franken EA, Jr., Dorfman DD *et al.*: "Role of faulty decision making in the satisfaction of search effect in chest radiography". *Academic Radiology.* **7**: 1098-106, 2000.
51. Samuel S, Kundel HL, Nodine CF *et al.*: "Mechanism of satisfaction of search: Eye position recordings in the reading of chest radiographs". *Radiology.* **194**: 895-902, 1995.
52. Nodine CF and Kundel HL: "Using eye movements to study visual search and to improve tumor detection". *RadioGraphics.* **7**: 1241-1250, 1987.
53. Kundel HL, Nodine CF and Krupinski EA: "Computer displayed eye position as a visual aid to pulmonary nodule interpretation". *Investigative Radiology.* **25**: 890-896, 1990.
54. Nodine CF, Kundel HL, Mello-Thoms C *et al.*: Role of computer assisted visual search in mammographic interpretation, in Krupinski E andChakraborty D: *Image Perception and Performance,* Proc SPIE, 2001, vol. 4324, pp 52-55.
55. Nodine CF, Kundel HK, Mello-Thoms C *et al.*: "How experience and training influence mammography expertise". *Academic Radiology.* **6**: 575-585, 1999.
56. Christensen EE, Murry RC, Holland K *et al.*: "The effect of search time on perception". *Radiology.* **138**: 361-365, 1981.
57. Nodine CF, Mello-Thoms C, Kundel HL *et al.*: "The time course of perception and decision making during mammographic interpretation." *AJR American Journal of Roentgenology*: 917-923, 2002.
58. Carmody DP, Kundel HL and Toto LC: "Comparison scans while reading chest images: Taught, but not practiced". *Investigative Radiology.* **19**: 462-466, 1984.
59. Nodine CF, Mello-Thoms C, Weinstein SP *et al.*: "Blinded review of retrospectively visible unreported breast cancers: an eye-position analysis". *Radiology.* **221**: 122-129, 2001.

60. Mello-Thoms C, Dunn SM, Nodine CF *et al.*: "The perception of breast cancers - A spatial frequency analysis of what differentiates masses from reported cancers." *IEEE Transactions on Medical Imaging.* **22**: 1297-1306, 2003.
61. Moise A and Atkins SM: "Radiology workstation design for the medical intensive care unit". *J Digit Imaging.* **15 Suppl 1**: 151-5, 2002.

Acknowledgements:

The work that is reviewed here was supported for many years by Grant CA-32870 from the National Cancer Institute, NIH, USPHS and was carried out with many colleagues. Among my co-investigators were Calvin Nodine, George Revesz, Lawrence Toto, Dennis Carmody, Elizabeth Krupinski, Susan Weinstein, and Claudia Mello-Thoms. There were many other colleagues, radiologists and radiology residents who contributed criticism and many tedious hours of viewing time.