How To Minimize Perceptual Error and Maximize Expertise in Medical Imaging.

Harold L. Kundel
Department of Radiology
University of Pennsylvania Health Sciences Center
3600 Market Street, Suite 370, Philadelphia, PA, 19444

ABSTRACT

Visual perception is such an intimate part of human experience that we assume that it is entirely accurate. Yet, perception accounts for about half of the errors made by radiologists using adequate imaging technology. The true incidence of errors that directly affect patient well being is not known but it is probably at the lower end of the reported values of 3 to 25 %. Errors in screening for lung and breast cancer are somewhat better characterized than errors in routine diagnosis. About 25% of cancers actually recorded on the images are missed and cancer is falsely reported in about 5% of normal people.

Radiologists must strive to decrease error not only because of the potential impact on patient care but also because substantial variation among observers undermines confidence in the reliability of imaging diagnosis. Observer variation also has a major impact on technology evaluation because the variation between observers is frequently greater than the difference in the technologies being evaluated. This has become particularly important in the evaluation of computer aided diagnosis (CAD).

Understanding the basic principles that govern the perception of medical images can provide a rational basis for making recommendations for minimizing perceptual error. It is convenient to organize thinking about perceptual error into five steps. 1) The initial acquisition of the image by the eye-brain (contrast and detail perception). 2) The organization of the retinal image into logical components to produce a literal perception (bottom-up, global, holistic). 3) Conversion of the literal perception into a preferred perception by resolving ambiguities in the literal perception (top-down, simulation, synthesis). 4) Selective visual scanning to acquire details that update the preferred perception. 5) Apply decision criteria to the preferred perception.

The five steps are illustrated with examples from radiology with suggestions for minimizing error. The role of perceptual learning in the development of expertise is also considered.

Keywords: Keywords: Human error, visual perception, cancer screening, radiology, expertise

1. ERROR AND VARIATION IN IMAGING DIAGNOSIS

In a delightful monograph summarized his own experience and tabulated many studies of error and variation done up to 1967 Smith (Smith 1967) defined error in a broad sense as a "wandering from the truth" based upon its Latin origin in the word errare, to go astray. Rigorous analysis requires a more specific definition and biostatisticians generally describe accuracy rather than error apparently preferring descriptions of success rather than failure. Accuracy, which is the likelihood that a reader is correct with respect to an established reference standard, is described either as a sensitivity-specificity pair or as a parameter of the receiver operating characteristic (ROC) curve. Sensitivity is the probability of reporting a diseased person as diseased and specificity is the probability of reporting a disease-free person as disease-free. The ROC curve, which is a plot of the true positive fraction (sensitivity) against the false positive fraction (1- specificity), shows all of the possible sensitivity-specificity pairs as decision criteria are varied from strict to lenient (see Figure 1).

Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment,
edited by Yulei Jiang, Berkman Sahiner, Proc. of SPIE Vol. 6515
651508, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.718061

Agreement is the likelihood of one reader giving the same response as another reader and a lack of agreement among readers produces variation. The large variability in image reader performance has been known since the classic 1947 study of tuberculosis screening failed to show which of four chest imaging systems was best because the differences among observers was greater than the differences among the imaging systems(Birkelo et al. 1947). More recently, variability in mammogram and chest interpretation has been well documented using image test sets and relatively large number of readers. Elmore et al.(Elmore et al. 1994) studied 10 radiologists reading 150 mammograms and found a median diagnostic inconsistency of 22%. Beam et al.(Beam et al. 1996) studied a random sample of 108 radiologists from mammography screening centers who interpreted the same set of 79 screening mammograms. They found a range of about 40% in screening sensitivity and a variability of 11% in the ability to detect cancer. Potchen et al. (Potchen et al. 2000) presented a standardized set of 60 chest radiographs to 162 participants with different levels of training. They found a range of 16% in the average area under the ROC curve of the 20 best and 20 worst radiologists.

Given the repeated documentation of performance variability over the last three decades, it should make sense to study methods for decreasing reader variability, yet the radiological community always seems to be in denial about the magnitude and the importance of reader error and variation. Admittedly, the true incidence of errors that directly affect patient well being is not known. In a survey of the accuracy of chest radiography, computed tomography, barium studies, and MRI, Goddard et al. (Goddard et al. 2001) estimated that clinically significant errors ranged from 2 to 20%. Errors in screening for lung and breast cancer using standard radiography are somewhat better characterized than errors in general diagnosis. The National Cancer Institute sponsored a lung cancer screening study in three institutions which enrolled 31,293 male volunteers, over age 45 who smoked at least one pack of cigarettes a day (Berlin et al. 1984). About 50% of the cancers were missed and about 2% were falsely reported. In a multi-institutional study of 49,528 asymptomatic women an average of 60% of the cancers were missed and 2% were falsely reported (Pisano et al. 2005). ROC curves and the "operating points" for the two studies are shown in Figure 1.

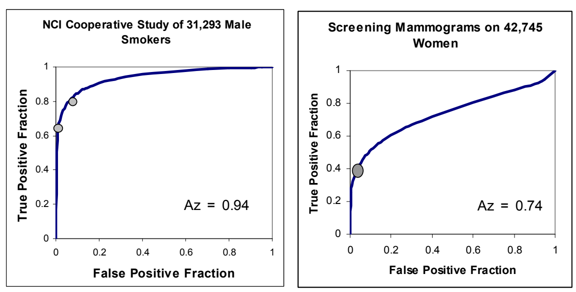

Fig 1 . ROC curves for cancer screening.
The "operating points" in the NCI screening study are for definitely cancer and probably cancer.
The single "operating point" in the mammogram screening study is the reported sensitivity and specificity pair.

2. VISUAL IMAGING SYSTEMS

Almost every medical imaging systems in common use can be characterized as a *visual imaging system* because a human being is required to render a diagnostic decision. Morgan (Morgan 1966) stated that every visual imaging system has four components:
(1) a patient or some other object of interest,
(2) a source of radiant energy that interacts with the object and emits a radiant signal,
(3) a device that captures the signal and converts it into a form suitable for analysis,
(4) an intelligence that analyzes the converted signal and gives a report about the object.

It is important to emphasize that the output of the system is not an image but an interpretation of the signal in the form of a report or a decision about the object (Kundel 1979). The analytic intelligence could be programmed into a computer eliminating the need for a visually accessible image but for the foreseeable future the image form suitable for analysis will continue to be a visual display and the intelligence will be provided by human readers.

Any discrepancy between the true state of the object and the reported state of the object is a combination of the uncertainty (noise) that is introduced at each step in the transformation of the signal from component to component and variation in the application of decision criteria by the reader. Most investigators divide errors into those that that originate before and after the production of the displayed image or more formally those attributable to technological factors and those attributable to human perception and judgment. Using this distinction Renfrew et al.(Renfrew et al. 1992) estimate that about 60% of the errors in 182 cases that they reviewed were attributable to perception and judgment. The use of ROC analysis sometimes can separate perceptual and judgmental errors and cleverly designed experiments can probe subsets of perceptual error such as visual search and pattern recognition (Kundel 2004).

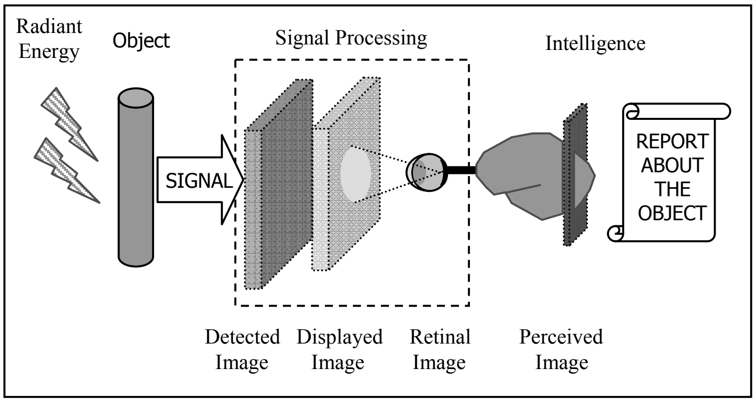

Fig 2. A visual imaging system

The diagram shows the principal images that exist at different stages as the signal is passed through the system. It also shows that the output is a report about the object.

3. MINIMIZING ERRORS

3.1. THE IMAGE ACQUISITION STAGE

3.1.1. Visual Input - Matching the Eye-Brain to the Display

Digital imaging technology enables the viewer to match the display to the eye-brain by normalizing the display grayscale (Blume and Hemminger 1997) and providing controls to adjust the contrast range (called window and level). However, even after the primary adjustment of the display, there is more that can be done to match the eye-brain to the display.

3.1.2. Viewing Distance and Image Size - The Visual System Transfer Function

The displayed image that is projected onto the retina is modified by the optics of the eye. Additional modification occurs in the retinal neural network. The result can be described by the eye-brain modulation transfer function (MTF), which shows how the contrast of sine waves of different spatial frequencies is modulated by the optics of the eye and the sensory network of the retina. Since it is not strictly an optical transfer function some authors prefer to call it the visual system transfer function.

The MTF of the eye-brain plotted using the model of Barten (Barten 1999) is shown in figure 3. The usual plot of contrast sensitivity vs. spatial frequency is shown on the left while the inverse of contrast sensitivity, minimal detectable contrast, is shown on the right. To state the obvious, small object are hard to see. It is less obvious that large objects or even small objects viewed from up close may also be harder to see. Shea and Ziskin showed that small pulmonary nodules are missed when the reader is too close (Shea and Ziskin 1972). This is a direct consequence of the low frequency rolloff in the eye-brain MTF. Another consequence is that Mach bands appear at many boundaries in radiographic images (Edholm 1981). These light or dark bands that appear wherever there are density gradients are not physically present in the image but are created in the first stage of processing by the neural network of the retina. Sometimes they can be mistaken for pathology (Daffner 1983).

Every structure in an image is not visualized with optimal contrast at a single viewing distance. Contrast is optimal when a boundary or a feature in an object subtends a visual angle of three to five degrees.

It is obvious that magnifying optimizes the visualization of small objects but it is not as obvious that minifying optimizes the visualization of large objects.

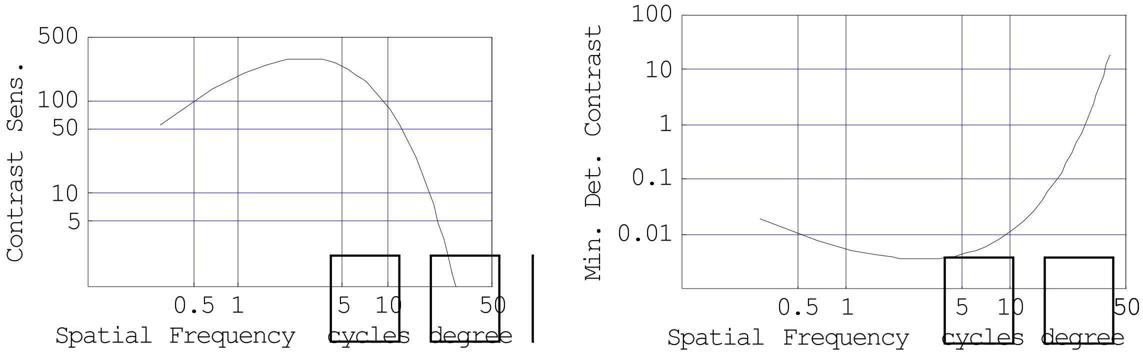

Fig 3. The Visual System Transfer Function (Eye-Brain MTF)

3.1.3. The Effect of Luminance on Contrast Thresholds

Measurement of absolute contrast thresholds requires that the eye is fully adapted to the luminance of the uniform background. Then the task is the detection of a single, small spot of light superimposed onto the uniform background. The best data is that of Blackwell (Blackwell 1946), which is well summarized in the paper by Robson et al. (Robson et al. 1996). At low light levels (below 1 cd/m^2) the contrast threshold is photon limited and changes as the square root of the luminance. This is called the Rose-DeVries zone because Rose and DeVries independently developed the photon limited hypothesis. At high light levels the contrast threshold is constant. This is called the Weber-Fechner zone because Weber first observed the constant threshold and Fechner formulated it mathematically.

The functional contrast threshold requires discrimination of contrast between two objects that are near each other but not necessarily at the luminance of the background. The eye is adapted to the combined average luminance of the background and the luminance of the object of interest, which is in the center of the visual field. The data of Heinemann (Heinemann 1961) show that there is a family of curves that describe the relationship between the luminance that induces adaptation and contrast sensitivity. The functional contrast threshold depends upon the instantaneous level of light adaptation in the retina. Figure 4 shows the some idealized curves as well as the arrangement of the stimulus used to collect data for both the absolute and the functional threshold (Schreiber 1986). The curves superimposed on the absolute threshold curve, which are somewhat exaggerated, show the results when the background is dim (10 cd/m^2) and bright (100 cd/ m^2). The optimal contrast sensitivity occurs when the luminance of the test field is the same as the luminance of the background.

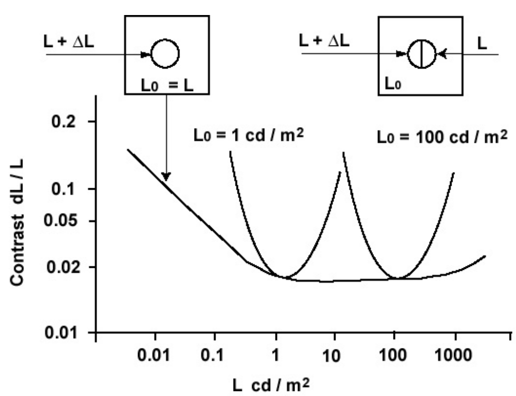

Fig. 4. The effect of luminance on contrast sensitivity The arrangement of the test field shown on the left is used to measure the absolute contrast sensitivity. The test field arrangement shown on the right is closer to a real image where the average luminance may vary widely from place to place.

In images with large bright areas like the heart and mediastinum in a chest image the visual adaptation level may not be optimal especially for seeing objects in the dark areas of the lungs. Inverting or masking the bright areas or viewing a small window against a dark background can usefully optimize contrast.

3.1.4. Glare and Ambient Illumination

Glare is due to relatively bright light either emitted or reflected either from the display or from a light source outside the display. The analogy in radiology is the light around the film on an illuminator or the white border on a display. Glare changes the functional light adaptation. Instead of adapting to the average luminance of the displayed image the eye adapts to the brighter extraneous light and loses contrast sensitivity for the darker areas.

Ambient light is the light scattered in the surroundings. It has a direct and indirect effect on the contrast threshold. First, it directly changes the light adaptation of the eye and, second, light reflected from the display increases the scattered light and decreases the displayed contrast.

It should be apparent that sources of glare should be removed from the viewing environment and the level of ambient illumination should be appropriate for the luminance of the images on the workstation. Personal observations of the author suggest that this is the most frequent breach of common sense that can lead to errors of detection.

3.1.5. Selected Studies of Contrast Sensitivity in Radiology

Robson et al. (Robson, Kotre et al. 1996) studied the effect of lightbox luminance (high luminance levels) on the detection of low contrast objects on mammograms. They found a constant contrast sensitivity, which would be expected from Blackwell's (Blackwell 1946) data. There also are a studies of the effect of ambient light on the detection of low contrast objects on illuminators (Alter et al. 1982), microcalcifications on illuminators (high luminance levels) (Kimme-Smith et al. 1997), and low contrast objects on CRT's (low luminance levels) (Rogers et al. 1987; Itoh et al. 1992; Krupinski and Roehrig 1998). The data are consistent with the proposition that low light levels decrease functional contrast sensitivity and that ambient light levels decrease functional contrast sensitivity. The data of Itoh et al.(Itoh, Ishigaki et al. 1992) superficially seem contradictory but they show that performance is best when the ambient light matches the display luminance. The data indicate that a room can actually be too dark for optimal viewing of a display. All of the variables are very simply defined in a review by Dwyer et al. (Dwyer III et al. 1992).

3.1.6. Lessons for the Reader

Do not pretend that we know nothing about contrast and detail perception.
Adjust the lights.
Eliminate glare and reflection.
Use the zoom for magnification and minification.

3.2. Image Organization – The Literal Perception

3.2.1. Developing a preferred perception – The influence of prompts

Once the displayed image is transferred to the brain it must be converted into a logical perception. Rock (Rock 1983) proposed that perception resulted from two activities, the development of a literal perception and modification of the literal perception to form a preferred perception. The literal perception is free of inferences about the properties of the object that was imaged. It is rapid and automatic and has been called bottom-up processing. The preferred perception is a modification of the literal perception that takes into account the visual task and prior knowledge about the object. It is a slower process that utilizes continuous visual input as the eyes are scanned over the image. It has been called top-down processing.

Kundel and Nodine (Kundel and Nodine 1983) showed a group of laypersons and radiologists a puzzle picture and a series of radiological images and they were asked to draw what they saw. Figure 5 shows the ultrasound image of a carcinoma of the head of the pancreas and the drawing of a layperson and a radiologist. The radiologist “knew” what the image represented and the drawing was somewhat idealized. Notice that a boundary indicating the inferior part of the mass is drawn where none exists in the image. Tuddenham (Tuddenham 1962) cites examples of contours that do not exist in the image that are, nevertheless, perceived as being present because the viewer knew that they must exist in the patient. Figure 6 shows the results of viewing a puzzle picture of the head of a cow before and after the viewer was made aware of the content of the image. The “before” drawing, although not exactly what Rock meant by the literal perception, illustrates the concept of organizing an image without any preconceived idea about the content of the image. Once the person was told about the cow, the image was reorganized.

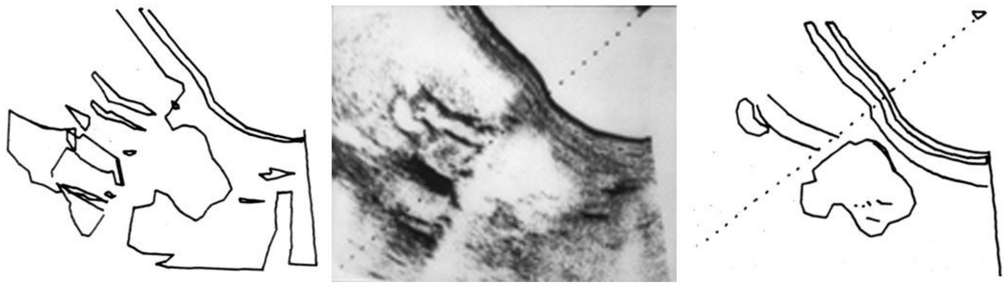

Fig. 5. An ultrasound image of a carcinoma of the head of the pancreas (center) and a sketch by someone with (right) and without (left) knowledge about the meaning of the picture.

Prior knowledge influences perception. In medical imaging prior knowledge starts with anatomy, pathology, and familiarity with the physical principles that govern the formation of images. This knowledge should be used by every competent radiologist. It is harder to determine unambiguously the effect of prompts that in effect tell the reader what to expect in the image, much as the prompt “this is a picture of the head of a cow” changed the perception of the puzzle picture. In radiology prompts are given by the clinical history supplied with the image. Some of the published studies show that clinical history helps (Aideyan et al. 1995) others that it is neutral (Good et al. 1990; Elmore et al. 1997)and yet others that it can actually degrade performance (Eldevik et al. 1982) The individual studies are strongly biased by the choice of test cases and readers and by the experimental design. Renfrew et al. (Renfrew, Franken Jr. et al. 1992) attributed 10% of the 182 errors in their series to poor communication. Clinical history may be wrong or misleading but it is foolhardy for the reader to ignore it.

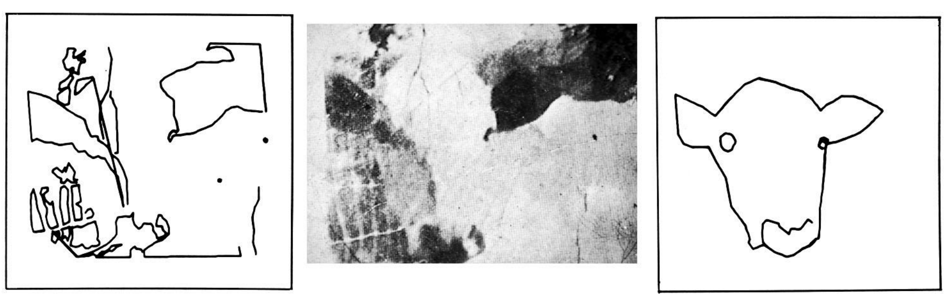

Fig. 6. A puzzle picture of the head of a cow (center) and a sketch by someone who did not see the cow (left) and who saw the cow after a prompt (right).

3.2.2. Lessons for the Reader

Look at the images without reading the clinical history and come to a conclusion.
Read the clinical history, look again, and modify the conclusion if necessary.

3.3. Image Organization – The Preferred Perception

3.3.1. Visual Search, Viewing Time, and Holistic Recognition

In the days of film radiology, an internist friend once asked me why a certain radiologist on our staff frequently made the diagnosis before the films were completely on the lightbox. This prompted a study of "flash viewing", in which ten radiologists were asked to make a diagnosis in 200 milliseconds (Kundel and Nodine 1975). A variety of cardiac configurations as well as pneumonia, a pneumothorax, a lung mass, metastatic nodules diffuse small nodules and a solitary lung nodule were shown. The radiologists had 97% true positives when given unlimited time and 70% true positives when shown the images for 200 milliseconds. Oestmann (Oestmann et al. 1988) did a flash viewing study using pulmonary nodules and showed that the true positive fraction for subtle and obvious lung cancers was 74% and 98% in unlimited viewing time and 30% and 70% when viewed for 250 milliseconds. Furthermore, the detection in four seconds was statistically undistinguishable from unlimited viewing time.

Clearly much of what happens in perception precedes exhaustive visual scanning of the image. Recordings of the location of the initial few fixations of a group of nine mammographers and mammography trainees showed that on about half of the images the mammographers jumped right to the cancer whereas most of the trainees only jumped to the cancer in 20% of the images (Kundel et al. 2007). This is presented as evidence that the search for cancers in mammograms begins with an holistic perception that shows the perceptual system where to look for cancers. The most proficient readers develop the ability to use the holistic perception for search guidance while the less proficient readers must still depend on scanning the eyes over the image.

3.4. After the Perception Comes the Judgment

3.4.1. Classification - The Interface of Perception and Judgment

Once a stable perception is achieved the reader must classify or name the pathology that has been identified. The problem of agreement on terminology – or was it agreement about what was perceived in the images – was studied by three prominent radiologists, R. Newell, W.E. Chamberlain, and L. Rigler, who tried to agree about the classification of the appearance of tuberculosis on chest radiograms and despite numerous attempts could not agree.(Newell et al. 1954; Yerushalmy 1969). Consistency in describing the appearance of diffuse opacities in the lungs is still a still a current problem for readers who must classify the extent and severity of the pneumoconioses (Hodous et al. 1991; Gitlin et al. 2004). The variation among observers has been greatly reduced but not eliminated by the introduction of proficiency exams for readers who wish to be certified (Morgan 1979; Wagner et al. 1992). A similar change has occurred with the introduction of standards for the production and reading of mammograms (Geller et al. 2006).

3.4.2. Decision Making

The final report, which is the output of the visual imaging system, is influenced by the reader's propensity to use strict or lenient criteria for making a diagnosis. A small vague opacity in the lung of a 25 year old may either be neglected or called to the attention of the referring physician for a future follow-up examination while an identical appearing opacity in the lung of a 65 year old smoker will trigger a workup for lung cancer. The bias to either report or neglect a particular lesion is quantified by the receiver operating characteristic (ROC) curve (see Fig 1). A lot of observer variability can be explained by differences in the application of decision criteria (Manning et al. 2004). The ROC curve provides a description of the variation but does not provide a means for minimizing the bias that influences the reports.

4. EXPERTISE

4.1. Definition of Expertise

Expertise is the capability to perform a task in a limited domain with exceptional results when compared to others capable of performing the same task. The characteristics of expertise are accuracy and speed as well as simplicity and efficiency.

Despite the rather high error rates that have been reported, studies comparing the accuracy of radiologists and trainees (Potchen, Cooper et al. 2000) or radiologists and non-radiologist physicians (Eng et al. 2000) show the superiority of the radiologist. The radiologists also make accurate diagnoses more rapidly than trainees (Christensen et al. 1981; Nodine et al. 2002).

The development of simplicity and efficiency can be seen in the visual scanpaths in Figure 7 that show the evolution of the search pattern from a first year medical student to a radiologist (Kundel and LaFollette 1972).

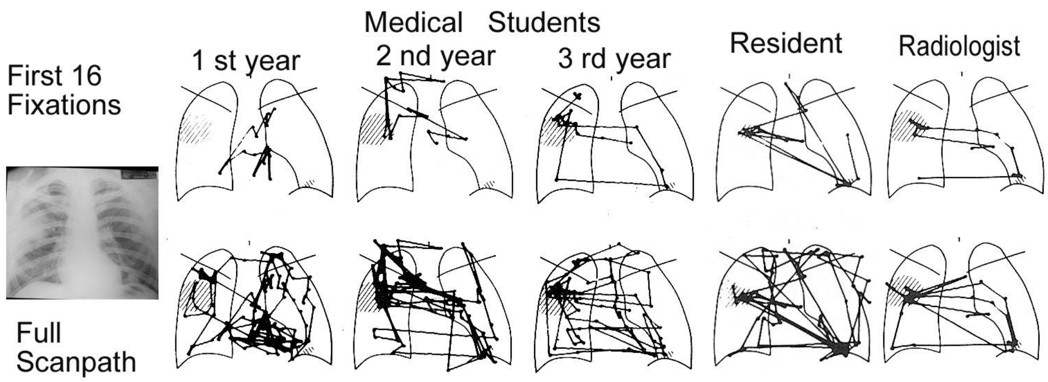

Fig. 7. Plots of the scanpath of persons of different levels of training from first year medical student to radiologist. They all had their eye-position recorded while looking at a chest with a large opacity in the right upper lobe and a smaller on in the left lower lobe. The upper set show the first 16 fixations (1.6 seconds of viewing) and the bottom set shows the completed scan.

The first year medical student scans the border of the right upper lobe opacity, does not scan the left lower lobe opacity but concentrates on the heart. The second year medical student scans both opacities ending up with a pattern very similar to the radiology resident but with many more fixations. The first year medical student had completed a course in anatomy where normal x-ray images were shown but had not been introduced to pathology. The second year student completed a course in pathology where abnormal x-ray images were shown. The third year medical student, the resident and the radiologist all scan both of the opacities in the first 16 fixations but the complete pattern of the radiologist is the simplest and the most efficient because it hits the relevant pathology quickly, broadly surveys the rest of the image, and stops.

4.2. The Role of Knowledge and Experience

The effect of knowledge and experience on performance has not been extensively studied in radiology. Herman and Hessel (Herman and Hessel 1975) did not find any difference between residents and staff radiologists interpreting chest images. They assumed that the residents had adequate knowledge and that the experience of the staff made no difference. This is lillustrated in Figure 7 where despite the differences in the scanpaths, all of the readers from second year resident and upward reported the images as abnormal and the third year medical student correctly gave a diagnosis of pneumonia. The increased knowledge and experience of the radiologist was of little help for this image. This may not be true for cancer detection on mammograms. Nodine et al.(Nodine et al. 1999) reported that the detection of cancer on mammograms improved with training (technologists, residents, mammographers) and with experience expressed as the number of cases read in the last year. Esserman et al.(Esserman et al. 2002) found a statistically significant difference between the performance of high and low volume mammography centers and Kan et al. (Kan et al. 2000) found statistically significant difference in cancer detection between readers who reported less than 2000 mammograms a year and readers who reported more.

4.3. Perceptual Learning

If we consider that there are two causes of variability in image interpretation. First, readers perceive the image differently and come to different conclusions and, second, although the perception is the same the criteria used to classify the images are applied differently. Much of the teaching of trainees and continuing education in radiology is aimed at reducing variability by getting everyone to agree about the classification of abnormalities. Less attention is given to seeking uniformity in perception. Very little research has been done on perceptual learning in radiology. It is not clear if it is possible to teach people to perceive things differently. Sowden et al. (Sowden et al. 2000) using dots to simulate microcalcifications in mammograms showed that people can indeed increase their sensitivity without changing their decision criteria. Kundel et al. (Kundel, Nodine et al. 2007) have shown that mammographers respond more rapidly to cancers in images than most trainees and have speculated that experience has caused a shift in the perceptual process from a primarily search mode to a holistic recognition mode. Presently, perceptual learning apparently occurs as a result of extensive experience looking at images, most of the time without any feedback about whether the interpretation is correct. Kundel and Nodine (Kundel and Nodine 1983) have suggested that “systematic visual training of students of radiology using typical examples of abnormality and patterns of normal variants” might be a useful approach to providing perceptual learning and reducing error and variation.

REFERENCES

Aideyan, U. O., Berbaum, K. and Smith, W. L. (1995). "Influence of prior radiologic information on the interpretation of radiographic examinations." Acad Radiol 2: 205-208.

Alter, A. J., Kargas, G. A., Kargas, S. A., Cameron, J. R. and McDermott, J. C. (1982). "The influence of ambient and viewbox light upon visual detection of low-contrast targets in a radiograph." Invest Radiol 17: 402-406.

Barten, P. G. J. (1999). Contrast Sensitivity of the Human Eye and Its Effect on Image Quality, SPIE-The International Society for Optical Engineering.

Beam, C. A., Layde, P. M. and Sullivan, D. C. (1996). "Variability in the interpretation of screening mammograms by US radiologists." Arch Intern Med 156: 209-213.

Berlin, N. I., Buncher, C. R., Fontana, R. S., Frost, J. K. and Melamed, M. R. (1984). "The National Cancer Institute cooperative early lung cancer detection program." Am Rev Resp Dis 130: 545-549.

Birkelo, C. C., Chamberlain, W. E., Phelps, P. S., Schools, P. E., Zacks, D. and Yerushalmy, J. (1947). "Tuberculosis case finding. A comparison of the effectiveness of various roentgenographic and photofluorographic methods." JAMA 133: 359-366.

Blackwell, H. R. (1946). "Contrast thresholds of the human eye." JOSA 36: 624-643.

Blume, H. and Hemminger, B. M. (1997). "Image presentation in digital radiology: Perspectives on the emerging DICOM display function standard and its application." Radiographics 17: 769-777.

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

Daffner, R. H. (1983). "Visual illusions affecting perception of the roentgen image." CRC Critical Reviews in Diagnostic Imaging 20: 79-119.

Dwyer III, S. J., Steward, B. K., Sayre, J. W., Aberle, D. R., Boechat, M. I., Honeyman, J. C., Boehme II, J. M., Roehrig, H., Ji, T.-L. and Blaine, G. J. (1992). "Performance characteristics and image fidelity of gray-scale monitors." Radiographics 12: 765-772.

Edholm, P. (1981). "Boundaries in the radiographic image." Acta Radiol Diag 22: 457-473.

Eldevik, O. P., Dugstad, G., Orrison, W. W. and Haughton, V. M. (1982). "The effect of clinical bias on the interpretation of myelography and spinal computed tomography." Radiology 145: 85-89.

Elmore, J. G., Wells, C. K., Howard, D. H. and Feinstein, A. R. (1997). "The impact of clinical history on mammographic interpretations." JAMA 277: 49-52.

Elmore, J. G., Wells, C. K., Lee, C. H., Howard, D. H. and Feinstein, A. R. (1994). "Variability in radiologists' interpretation of mammograms." NEJM 331: 1493-1499.
Eng, J., Mysko, W. K., Weller, G. E. R., Renard, R., Gitlin, J. N., Bluemke, D. A., Magid, D., Kelen, G. D. and Scott, J., W.W. (2000). "Interpretation of emergency department radiographs: a comparison of emergency medicine physicians with radiologists, residents with faculty, and with digital display." AJR Am J Roent 175: 1233-1238.
Esserman, L., Cowley, H., Eberle, C., Kirkpatrick, A., Chang, S., Berbaum, K. and Gale, A. (2002). "Improving the accuracy of mammography: volume and outcome relationships." Journal of the National Cancer Institute 94: 369-375.
Geller, B. M., Ichikawa, L. E., Buist, D. S., Sickles, E. A., Carney, P. A., Yankaskas, B. C., Dignan, M., Kerlikowske, K., Yabroff, K. R., Barlow, W. and Rosenberg, R. D. (2006). "Improving the concordance of mammography assessment and management recommendations." Radiology 241: 67-75.
Gitlin, J. N., Cook, L. L., Linton, O. W. and Garrett-Mayer, E. (2004). "Comparison of "B" readers' interpretations of chest radiographs for asbestos related changes." Acad Radiol 11: 843-856.
Goddard, P., Leslie, A., Jones, A., Wakeley, C. and Kabala, J. (2001). "Error in radiology." Br J Radiol 74: 949-951.
Good, B. C., Cooperstein, L. A., DeMarino, G. B., Miketic, L. M., Gennari, R. C., Rockette, H. E. and Gur, D. (1990). "Does knowledge of the clinical history affect the accuracy of chest radiograph interpretation?" AJR Am J Roent 154: 709-712.
Heinemann, E. (1961). "The relation of apparent brightness to the threshold for differences in luminance." Journal of Experimental Psychology 61: 389-399.
Herman, P. G. and Hessel, S. J. (1975). "Accuracy and its relationship to experience in the interpretation of chest radiographs." Invest Radiol 10: 62-67.
Hodous, T. K., Chen, R. A., Kinsley, K. B., Liu, X. T., McLaughlin, J. K., Chen, J. Q., Wu, Z. E. and Blot, W. J. (1991). "A comparison of pneumoconiosis interpretation between Chinese and American readers and classifications." J Tongji Med Univ 11: 225-229.
Itoh, Y., Ishigaki, T., Sakuma, S., Hirose, M., Fukatsu, H., Itoh, S., Horikawa, Y., Shimamoto, K., Tadakoro, T., Ikeda, M. and Itoh, K. (1992). "Influence of CRT workstation on observer's performance." Computer Methods and Programs in Biomedicine 37: 253-258.
Kan, L., Olivotto, I. A., Warren Burhenne, L. J., Sickles, E. A. and Coldman, A. J. (2000). "Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program." Radiology 215: 563-567.
Kimme-Smith, C., Haus, A. G., DeBruhl, N. and Bassett, L. H. (1997). "Effects of ambient light and view box luminance on the detection of calcifications in mammography." AJR Am J Roent 168: 775-778.
Krupinski, E. and Roehrig, H. (1998). Influence of monitor luminance and tone scale on observer detection performance. Medical Imaging 98: Image Perception, SPIE.
Kundel, H. (2004). Reader error, object recognition and visual search. Medical Imaging 2004. D. Chakraborty and M. Eckstein, SPIE. 5372: 1-11.
Kundel, H. L. (1979). "Images, image quality and observer performance." Radiology 132: 265-271.
Kundel, H. L. and LaFollette, P. S. (1972). "Visual search patterns and experience with radiological images." Radiology 103: 523-528.
Kundel, H. L. and Nodine, C. F. (1975). "Interpreting chest radiographs without visual search." Radiology 116: 527-532.
Kundel, H. L. and Nodine, C. F. (1983). "A visual concept shapes image perception." Radiology 146: 363-368.
Kundel, H. L., Nodine, C. F., Conant, E. F. and Weinstein, S. P. (2007). "Holistic component of image perception in mammogram interpretation: Gaze-tracking study." Radiology 242: 396-402.
Manning, D. J., Ethell, S. C. and Donovan, T. (2004). "Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph." British Journal of Radiology 77: 231-235.
Morgan, R. H. (1966). "Visual perception in fluoroscopy and radiography. Annual oration in memory of John D. Reeves, Jr., M. D., 1924-1964." Radiology 86: 403-416.
Morgan, R. H. (1979). "Proficiency examination of physicians for classifying pneumoconiosis chest films." AJR Am J Roentgenol 132: 803-808.
Newell, R. R., Chamberlain, W. E. and Rigler, L. (1954). "Descriptive classification of pulmonary shadows. A revelation of unreliability in the roentgen diagnosis of tuberculosis." Am Rev Tuberc 69: 566-584.
Nodine, C. F., Kundel, H. K., Mello-Thoms, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C. and Conant, E. F. (1999). "How experience and training influence mammography expertise." Acad Radiol 6: 575-585.

Nodine, C. F., Mello-Thoms, C., Kundel, H. L. and Weinstein, S. P. (2002). "The time course of perception and decision making during mammographic interpretation." AJR Am J Roent: 917-923.
Oestmann, J. W., Greene, R., Kushner, D. C., Bourgouin, P. M., Linetsky, L. and Llewellyn, H. J. (1988). "Lung lesions: Correlation between viewing time and detection." Radiology 166: 451-453.
Pisano, E. D., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, J. K., Acharyya, S., Conant, E. F., Fajardo, L. L., Bassett, L., D'Orsi, C., Jong, R. and Rebner, M. (2005). "Diagnostic performance of digital versus film mammography for breast-cancer screening." NEJM 353: 1773-1783.
Potchen, E. J., Cooper, T. G., Sierra, A. E., Aben, G. R., Potchen, M. J., Potter, M. G. and Siebert, J. E. (2000). "Measuring performance in chest radiography." Radiology 217: 456-459.
Renfrew, D. L., Franken Jr., E. A., Berbaum, K. S., Weigelt, F. H. and Abu-Yousef, M. (1992). "Error in radiology: Classification and lessons in 182 cases presented at a problem case conference." Radiology 183: 145-150.
Robson, K. J., Kotre, C. J. and Faulkner, K. (1996). "An experimental investigation of the effect of light-box luminance on the detection of low contrast objects in mammography." British Journal of Radiology 69: 153-159.
Rock, I. (1983). The Logic of Perception. Cambridge, MA, The MIT Press.
Rogers, D. C., Johnston, R. E. and Pizer, S. M. (1987). "Effect of ambient light on electronically displayed medical images as measured by luminance-discrimination thresholds." J Opt Soc Am 4: 976-983.
Schreiber, W. F. (1986). Fundamentals of Electronic Imaging Systems. Berlin, Springer-Verlag.
Shea, F. J. and Ziskin, M. C. (1972). "Visual system transfer function and optimal viewing distance for radiologists." Invest Radiol 7: 147-148.
Smith, M. J. (1967). Error and variation in diagnostic radiology. Springfield, Ill, Thomas.
Sowden, P. T., Davies, I. R. L. and Roling, P. (2000). "Perceptual learning of the detection of features in X-ray image: a functional rolse for improvements in adults' visual sensitivity?" Journal of Experimental Psychology 26: 379-390.
Tuddenham, W. J. (1962). "Visual search, image organization, and reader error in roentgen diagnosis." Radiology 78: 694-704.
Wagner, G. R., Attfield, M. D., Kennedy, R. D. and Parker, J. E. (1992). "The NIOSH B reader certification program. An update report." J Occup Med 34: 879-884.
Yerushalmy, J. (1969). The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. Radiol Clin N Am. W. J. Tuddenham. Philadelphia, W.B.Saunders. 7: 381-392.