History of Research in Medical Image Perception

Harold L. Kundel, MD

Human observers engage in 2 interrelated processes when interpreting medical images: perception and analysis. Perception is the unified awareness of the content of a displayed image that is present while the stimulus is on. Analysis is determining the meaning of the perception in the context of the medical problem that initiated the acquisition of the image. Radiologists have, correctly, regarded image analysis as their primary field of research. They have naively assumed that what they perceive in images is a faithful representation of the images' information content and have not been concerned with perception unless it fails. Failures have stimulated research on quantifying observer performance, defining image quality, and understanding perceptual error. This article traces the historical development of the use of receiver operating characteristic analysis for describing performance, the development of signal-to-noise ratio psychophysical models for defining task-dependent image quality, studies of error in small lesion detection, and the beginnings of studies of the nature of expertise in image interpretation. The history is traced through published articles.

Key Words: Perception research

J Am Coll Radiol 2006;3:402-408. Copyright © 2006 American College of Radiology

THE SCOPE OF PERCEPTUAL RESEARCH IN RADIOLOGY

Human observers engage in 2 interrelated processes when interpreting medical images: perception and analysis. Perception is defined as the unified awareness of the content of a displayed image that is present while the stimulus is on [1]. Analysis is determining the meaning of the perception in the context of the medical problem that initiated the acquisition of the image [2]. Radiologists have, correctly, regarded image analysis as their primary field of research. They have naively assumed that what they perceive in images is a faithful representation of the images' information content and have not been very concerned with the process of perception itself, until it fails. Failures show up as observer error and uncertainty, both of which affect judgments about image quality, attempts to objectively evaluate imaging technology, and especially everyday image interpretation.

Research on the perceptual component of image interpretation has largely but not exclusively focused on psychophysics, which is the study of the quantitative relationship between a visual stimulus and an observer's response. Although mainly descriptive, the ultimate goal of psychophysics is the development of mathematical models that allow the prediction of the system output from any arbitrary input [3]. That is, imaging scientists would like to be able to predict how an observer will respond to any image configuration without having to bother with the messy business of performing a study with real human observers [4]. Research in the broader domain of the mechanism of perception has mainly focused on understanding observer error.

This essay tracks the development of research in perception and psychophysics in radiology through publications, citing both the articles that introduced new ideas and those that summarized them. The original articles are not necessarily the best ones. Original work is frequently fuzzy; clarification comes later. There are innovators and popularizers in every human endeavor. Many of the central ideas have been summarized in invited lectures given at major radiology society meetings. They form the backbone of this survey because they provide a glimpse into what the community thought was important at the time. This survey is also heavily biased by my interest in using observer performance for evaluating imaging technology and in understanding the sources of reader error.

THE BEGINNING OF SERIOUS PERCEPTION RESEARCH IN THE 1940S

Confronted with the miracle of the roentgen ray, early workers had little time for conscious consideration of the miracle of the human eye.
—W. J. Tuddenham [5]

Department of Radiology, University of Pennsylvania, Philadelphia, Pa.

Corresponding author and reprints: Harold L. Kundel, MD, University of Pennsylvania, Department of Radiology, 3600 Market Street, Suite 370, Philadelphia, PA 19104; e-mail: kundelh@uphs.upenn.edu.

0091-2182/06/$32.00 • DOI 10.1016/j.jacr.2006.02.023

The earliest report of psychophysical research in radiology was published in 1899, barely 4 years after Roentgen's momentous discovery. Béclère [6] reported on experiments that he conducted on the sensitivity of the retina to the light of a fluoroscopic screen. He observed that it took 20 minutes to achieve maximal visual sensitivity, that sensitivity depended on the color of the light, and that dark adaptation was "absent in the fovea." He correctly related his observations to the then-newly developing knowledge of the physiology of retinal rods and cones and concluded correctly that complete dark adaptation was essential to being able to see details during fluoroscopy.

Béclère's [6] study was an isolated example of perceptual research. Undoubtedly there were others, but recognizable research in image perception was dormant for almost 50 years. In the early 1940s, 2 important events occurred, initiating formal research into perceptual psychophysics in radiology. In 1941, W. E. Chamberlain [7] of Temple University in Philadelphia was asked to give the Carman Lecture—"Fluoroscopes and Fluoroscopy"—at the annual meeting of the Radiological Society of North America (RSNA). This lecture rekindled interest in dark adaptation, visual acuity, and the limitation on image quality imposed by the quantum nature of radiation. In 1944, the US Public Health Service and the US Veterans Administration formed the Board of Roentgenology, which initiated a study to evaluate the effectiveness of various imaging techniques for detecting pulmonary tuberculosis [8,9]. Chamberlain was chairman of the board.

THE BOARD OF ROENTGENOLOGY INITIATES STUDIES OF TECHNOLOGY EVALUATION USING OBSERVER PERFORMANCE

By the 1940s, radiologic physicists were characterizing imaging systems in terms of contrast rendition, spatial resolution, and the radiation dose required to make an image. G.C.E. Burger of the Philips Company was one of the first radiologic physicists to recognize the importance of the mutual dependence of contrast and size as determining factors for the threshold perceptibility of the details in images. In the 1930s, he began characterizing images using contrast-size diagrams, which he called "perception curves." The data for the curves were obtained by measuring the threshold visibility of holes of decreasing diameter arranged horizontally in phantoms with vertically oriented steps of increasing thickness [10]. The contrast-size diagram is an expression of a psychophysical model [4]. However, in the 1940s, it apparently was not good enough to convincingly predict the relative merit of 4 competing imaging systems for detecting pulmonary tuberculosis, and the Board of Roentgenology decided to measure observer performance directly. The community was surprised by the results that were published in *JAMA* in 1947 [11]. Five expert observers interpreted chest images made with 4 techniques (35-mm photofluorograms, 4 × 10 inch stereophotofluorograms, 14 × 17 inch paper negatives, and conventional 14 × 17 inch celluloid films) on 1,256 individuals, and they could not show that any of the methods, "not even the 14 × 17 inch celluloid," was superior to the others. The main reasons for not being able to show any differences (if they really existed) were that the investigators could not establish the correct diagnoses by a method independent of the images and that the variation among the observers was greater than the differences among the techniques. An editorial titled "The Personal Equation in the Interpretation of a Chest Roentgenogram" [12] that accompanied the article expressed astonishment at the magnitude of observer disagreement and stated, "These discrepancies demand serious consideration." The problems of establishing truth and accounting for reader variation still vex us today. An article about the large observer variation in mammography published in 1994 [13] was also accompanied by an editorial that again emphasized the importance of the problem but gave no hints for a solution [14]. "It's deja vu all over again" [15].

The use of the term *personal equation* in the *JAMA* editorial linked observer variability in radiology to a long line of observer studies going back to observational astronomy. The term was coined around 1876, when the astronomer and mathematician F. W. Bessel found differences between his own readings of the transit time of stars across the meridian and those of 5 other astronomers. He tried to resolve the differences by calculating "personal equations" that adjusted each astronomer's readings to match his—an early attempt at quantitative psychophysics— but this proved unsatisfactory [16]. The problem of interastronomer variation was finally sidestepped by using an instrument, the chronograph, for measuring transit time. Perhaps computers will eventually be used to eliminate observer variability in diagnostic imaging. In a review of his work, "The Perceptibility of Details in Roentgen Examination of the Lung," Burger [10] showed "perception curves" for 5 individuals (his Figure 9), each different from the others.

Birkelo et al's [11] study on the effectiveness of various imaging techniques for detecting pulmonary tuberculosis started a flurry of investigations into human error in imaging diagnosis that were summarized by J. Yerushalmy, the biostatistician on the project [9]. L. H. Garland [8], a radiologist at Stanford University who also participated in the project, described it in his RSNA presidential address in 1948, titled "On the Scientific Evaluation of Diagnostic Procedures." Garland [8] enumerated 3 research objectives:

(1) to determine reliable methods for measuring the relative number of lesions missed by a reader,
(2) to study the probable reasons for missing lesions and their characteristics, and
(3) to investigate methods of interpretation that might lead to a reduction in the number of lesions missed.

GARLAND'S FIRST OBJECTIVE

The Development of Receiver Operating Characteristic (ROC) Analysis

L. Lusted, a radiologist who had worked on the studies stimulated by the results published by Birkelo et al [11] and was interested in applying the principles of formal logic to radiologic diagnosis [17], tackled the problem of properly describing performance. By the late 1950s the statistics of observer performance studies generally were presented in terms of sensitivity and specificity. Radiologists discussed the results in terms of underreading or false-negatives and overreading or false-positives, but at that time, the covariation of the 2 types of error was not appreciated. This important insight, originally developed by psychologists and systems engineers [18], was introduced into radiologic thinking (and perhaps into general medicine) around 1960 by Lusted [19], who summarized it in an RSNA Memorial Fund lecture titled "Logical Analysis in Roentgen Diagnosis." He introduced the statistical-decision-theory approach to the analysis of observer response data. The approach requires an observer not only to make the usual yes-or-no response about the presence of pathology in an image but also to give a confidence report about each decision. The fractions of true-positive and false-positive responses at each confidence level are plotted, and the statistical decision theory model is used to fit the experimental points to a smooth curve. The curve is called an ROC curve, and the curve-fitting model provides 2 very important parameters and their standard errors: the area under the curve and an index of detectability (*d*). The area under the curve is a single-valued parameter for performance that is free of bias due to the use of decision criteria (the predisposition to overread or underread) and reflects only the ability to separate normal from abnormal. It ranges from .5 for guessing to 1.0 for perfect performance. The index of detectability is a somewhat more difficult parameter to understand. Simply, it is the observer's signal-to-noise ratio (SNR) for the decision task and typically has a value between 0.5 and 3.0, although it has a range from zero to infinity.

The ROC model was mostly a laboratory tool until 1979, when J. Swets of Bolt, Beranak and Newman, a perceptual psychologist funded by a grant from the National Cancer Institute, assembled a group of psychologists, radiologists, and radiologic physicists who planned and conducted a study using ROC analysis comparing brain tumor detection by radionuclide scanning and computed tomography [20]. The study was the first demonstration of comparing imaging modalities in a clinical setting using ROC analysis. The ROC method is now widely used in radiology. The original study stimulated a lot of methodologic research in technology evaluation dealing with experimental design, curve fitting, and statistical analysis. The original curve-fitting algorithm developed in 1969 by Dorfman and Alf [21] at the University of Iowa has been modified and incorporated into many computer analysis programs by research groups at the University of Chicago, led by C. Metz, and at the University of Iowa, led by K. Berbaum. The state-of-the-art of ROC analysis was summarized in 4 review papers that were published in 1989 [22-25]. In 1992, Dorfman, Berbaum, and Metz [26] collaborated to produce an ROC analysis computer program that combines the statistical decision model with a classical analysis of variance. This so-called DBM approach has become the benchmark methodology for ROC analysis and in turn has stimulated a lot of the current research into ROC methodology.

The Development of SNR Psychophysical Models Designed to Predict Performance From Physical Measurements on Imaging Systems

The Rose–De Vries Psychophysical Model. It is safe to say that most radiologists working today have never performed fluoroscopy in a darkened room, viewing the patient in the dim, yellow-green light of a zinc-cadmium sulfide fluorescent screen. In the 1940s, when Chamberlain [7] started to work on his Carman Lecture, screen fluoroscopy was all that was available. The need for dark adaptation was well established, although in the lecture, he pointed out that some radiologists were either skeptical about its value or too impatient to bother with it. He prepared for the lecture by reviewing the fundamental work on dark adaptation of the physiologist S. Hecht [27] and had his colleague G. Henny, a radiologic physicist, perform fundamental measurements of the threshold visibility of details at screen fluoroscopy using the contrast-size phantoms developed by Burger. Chamberlain pointed out that the fluoroscopic screen could adequately display details that were visible in bright light but that could not be seen even by the fully dark adapted eye. A 1,000-fold increase in brightness was needed to shift the eyes from scotopic (rod) to photopic (cone) vision, and typical of the mind-set in radiology, Chamberlin suggested a technologic solution: the image intensifier. The image intensifier, developed by Coltman [28] in the late 1940s, first became commercially available in the early 1950s and completely replaced direct screen fluoroscopy, obviating the need for 20 minutes of dark adaptation, thereby depriving the radiologist of the opportunity of reading the morning newspaper before starting fluoroscopy.

The work of Chamberlain and Henny is significant not only because it brought the image intensifier to the attention of the radiology community but also because it teamed up a radiologist and physicist. It drew on knowledge of perceptual psychology and image evaluation and used original observations to support a solution to a practical problem. It was hoped that in addition to improving detail visibility, the image intensifier would lower the fluoroscopic radiation dose. This did not occur, because image noise (some price had to be paid for increased brightness) limited the visibility of details. In 1949, Sturm and Morgan [29] described the effect of noise on the threshold visibility of details in x-ray images using a mathematical model originally proposed by H. de Vries [30] and elaborated by A. Rose [31] of the RCA Sarnoff Laboratory. It is variously known as the "Rose model," the "Rose–De Vries model," or the "De Vries–Rose model," depending on whether one is from the engineering or the vision research community. It is a psychophysical model in which the physical image property is characterized by the SNR and the observer response is threshold visibility. Basically, the model asserts that an image, to be just recognizable, must have a SNR that exceeds some threshold value. Morgan's group at Johns Hopkins University eventually extended the model to include the physiologic optics of the human eye. In 1966, Morgan [32] summarized the work in an annual oration at the RSNA titled "Visual Perception in Fluoroscopy and Radiography." Although it expanded the Rose–De Vries model, it continued using threshold detectability as the observer's response in the psychophysical equation.

Task-Dependent Image Quality. In 1972, D. Goodenough, K. Rossmann, and Lusted [33], then at the University of Chicago, used ROC analysis to compare imaging techniques in the laboratory. It was becoming clear that optimizing image quality involved trade-offs between contrast rendition, spatial resolution, and noise. In fact, Rossmann and Wiley [34] had pointed out already that image quality could not be defined independently of the imaging task. The powerful idea of task-dependent image quality began to influence studies of psychophysics in radiology.

In 1979, Wagner et al [35] at the US Food and Drug Administration reformulated the SNR psychophysical model using the detection of a small faint object as the task, the index of detectability from ROC analysis as the observer response and defining the SNR in terms of the system modulation transfer function, the noise power spectrum, and the size, contrast, and profile of the signal. Over the next 5 years, the model was applied to virtually all imaging systems that existed at the time [36].

Structured Noise: The Fly in the Ointment. One of the difficulties with the Rose–De Vries SNR models is that in real images, signals such as lung nodules or masses on mammograms are embedded in an anatomic background that acts as camouflage, blocking the perception of the lesion. The noise in the SNR formulation is considered to be random, whereas the camouflaging background has recognizable structure, such as ribs and blood vessels, that is not random but still affects detection. In 1972, Revesz et al [37] tried to quantify what they called the "structured" noise in the background and define a SNR for lesion conspicuity rather than lesion detection. Although the camouflaging effect of image structure has been verified [38], the incorporation into SNR models has been difficult to accomplish.

The Theory of the Ideal Observer. Statistical decision theory defines an ideal observer as one who "makes the best possible use of all information to reach a decision" [4]. The theory describes procedures for calculating the performance of the ideal observer. Burgess et al [39] showed that by comparing the response of the human observer and the ideal observer, the efficiency of an imaging decision task could be determined. This comparison provided insight into the amount of improvement in human performance that was possible and provided a method for comparing different imaging tasks. During the 1980s and 1990s, the laboratory of H. Barrett at the University of Arizona was very productive in the development of imaging psychophysics [40]. An up-to-date account of psychophysical models for visual detection, largely the work of Barrett's students, can be found in the chapters written by K. Myers [41] and M. Eckstein, C. Abbey, and F. Bochud [42] in *The Handbook of Medical Imaging*.

GARLAND'S SECOND OBJECTIVE

Imaging psychophysics concentrates on building mathematical models that will predict performance given the physical parameters of an imaging system. A researcher accepts the proposition that performance is inherently inaccurate and incorporates the error into the analysis. The root cause of error is not addressed because of the complexity of the human perceptual apparatus and the difficulty of performing meaningful experiments. Nevertheless, a few investigators in radiology began to probe some of the fundamental mechanism of perception as applied to medical imaging.

In 1962, William Tuddenham [43] of the University of Pennsylvania, presented the RSNA Memorial Fund lecture, titled "Visual Search, Image Organization, and Reader Error in Roentgen Diagnosis." He had previously written about the impact of retinal anatomy and physiology on contrast perception [44] and had done experiments on the visual search of radiographs [45]. His work marked the beginning of formal research into the mechanism of human visual perception as it applies to radiologic imagery. In 1969, Tuddenham [46] edited an issue of *Radiological Clinics of North America* that brought together authors from disciplines that either contributed ideas to or benefited from research in medical image perception. The slim volume contained articles on perceptual psychology [47], statistics [9], search behavior [48], image quality [49], image processing [50,51], computer diagnosis [52], and learning radiology [53].

Studies of Visual Search

One possible source of error had been pointed out by Tuddenham [54], who proposed that when an observer was satisfied with the meaning of an image, active search was stopped. Smith [55], in a delightful, anecdotal classification of observer errors, coined the term *satisfaction of search*. The phenomenon is real: observers do not report unexpected findings on images when they have found something suggested by the original search task [56,57]. Subsequent research using gaze tracking has shown that the unreported lesions actually are looked at but are disregarded [58,59]. Tuddenham's original notion of "satisfaction of meaning" is probably more descriptive of the actual phenomenon, but we are stuck with the catchy *satisfaction of search*.

Gaze tracking has also been used to study search for lung nodules [60], fractures [61], and cancers in mammograms [62]. In all instances, most unreported abnormalities were selected for attention by the gaze but apparently not recognized. In fact, the observation that many of them received prolonged gaze dwell time [63,64] has stimulated research on using feedback from gaze tracking as an aid to lung nodule and mammogram mass detection [65].

GARLAND'S THIRD OBJECTIVE

Tuddenham [66] concluded his volume on the perception of roentgen images with some personal reflections. He wrote,

> The ultimate solution to the problem of 'reader error' is not yet clear. It may lie in the further development of automated pattern recognition systems.. . .For the moment, however, it appears to me more probably to lie in the elucidation of consistent logical systems of film analysis with which to guide the perceptual learning of the radiologist and his paramedical assistants.

The trend in radiology has been to technologic solutions, with the development of computer-assisted diagnosis systems [67] and attempts to improve display technology [68]. As usual, the perceptual side has been neglected, although recently there has been interest in perceptual learning [69] and the development of expertise in imaging tasks [70-72].

The Growth of Medical Image Perception as a Distinct Discipline

The diversity of investigators—radiologists, psychologists, physicists, engineers, and statisticians—interested in medical image perception was an obstacle to any type of organized activity for exchanging ideas. People attended different meetings and belonged to different professional societies. In 1983, a group of radiologists, psychologists, and physicists interested in perception organized a conference held in Park City, Utah, called The Far West Image Perception Conference. The attendees liked the conference and decided to organize a second in 2 years. Thus began an ongoing 2-year cycle of spontaneously organized conferences with no sponsoring organization. Interest in medical image perception was the glue that held the group together and resulted in 8 conferences labeled "Far West," even though some of them were held on the East Coast. The name was finally changed to the Medical Image Perception Conference for the ninth and subsequent conferences. In 1994, a conference on image perception was added to the annual International Society for Optical Engineering Medical Imaging Conference, giving people interested in perception another forum for the exchange of ideas.

In 1996, the perception group organized the Medical Image Perception Society to promote medical image perception research and its application. The society also began to formally sponsor the Medical Image Perception Conference. It is an international society, and in 2005 the Medical Image Perception Society XI Conference was held in Windermere, England.

SOME PERSONAL REFLECTIONS

Recently, there has been a great deal of interest in the assessment of computer-assisted imaging systems. In a review of contemporary assessment methods subtitled "Lessons From Recent Experience," Wagner et al [73] stated that

> funding agencies and researchers work years to discover ways to improve mean performance for some modalities by something on the order of 0.05 points (in terms of ROC area), for example. These improvements can be readily masked by the contemporary level of reader variability.

"Reader variability" has been a major problem since the study of Birkelo et al [11]. It was elegantly demonstrated for mammography by Beam et al [74], who used the ROC methodology on a sample of 108 radiologists interpreting mammograms. The variability was not only in the application of diagnostic criteria but also in the absolute ability to detect and recognize abnormalities. The history of observer performance suggests that this is a problem that is not going to go away. It may be sidestepped with limited applications by human-free computer image analysis, but we may have reached the limits of Garland's first goal of measuring performance. It may be time to put more effort and resources into the second and third goals of understanding the reasons for errors and ways of teaching radiologists to perform better and more consistently. Research in the deeper aspects of image perception and in the interface between perception and analysis may hold the key to the problem of error and variability.

REFERENCES

1. The Random House college dictionary. Rev ed. New York, NY: Random House; 1988.
2. Kundel HL. Images, image quality and observer performance. Radiology 1979;132:265-71.
3. Regan D. Human perception of objects. Sunderland, Mass: Sinauer Associates; 2000.
4. Burgess A. Image quality, the ideal observer, and human performance of radiologic detection tasks. Acad Radiol 1995;2:522-6.
5. Tuddenham WJ. Dark adaptation. In: Bruwer A, ed. Classic descriptions in roentgenology. Springfield, Ill: Charles C. Thomas; 1964. p 741-7.
6. Béclère A. A physiologic study of vision in fluoroscopic examinations. In: Bruwer A, ed. Classic descriptions in diagnostic roentgenology. Springfield, Ill: Charles C. Thomas; 1964.
7. Chamberlain WE. Fluoroscopes and fluoroscopy. Radiology 1942;38: 383-412.
8. Garland LH. On the scientific evaluation of diagnostic procedures. Radiology 1949;52:309-28.
9. Yerushalmy J. The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. Radiol Clin N Am 1969;7:381-92.
10. Burger GCE. The perceptibility of details in roentgen examinations of the lung. Acta Radiol Diag 1949;31:193-222.
11. Birkelo CC, Chamberlain WE, Phelps PS, et al. Tuberculosis case finding. A comparison of the effectiveness of various roentgenographic and photofluorographic methods. JAMA 1947;133:359-66.
12. The "personal equation" in the interpretation of a chest roentgenogram. JAMA 1947;133:399-400.
13. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretation of mammograms. N Engl J Med 1994;331: 1493-9.
14. The accuracy of mammographic interpretation. N Engl J Med 1994;331: 1521-2.
15. Berra Y, Garagiola J, Berra D. The Yogi book. New York, NY: Workman; 1998.
16. Stigler SM. The history of statistics. Cambridge, Mass: Harvard University Press; 1968.
17. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis. Science 1959;130:9-21.
18. Green DM, Swets JA. Signal detection theory and psychophysics. New York, NY: John Wiley; 1966.
19. Lusted LB. Logical analysis in roentgen diagnosis. Radiology 1960;74: 178-93.
20. Swets JA, Pickett RM, Whitehead SF, et al. Assessment of diagnostic technologies. Science 1979;205:753-9.
21. Dorfman D, Alf EJ. Maximum likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating method data. J Math Psych 1969;6:487-96.
22. Berbaum KS, Dorfman DD, Franken EA Jr. Measuring observer performance by ROC analysis: indications and complications. Invest Radiol 1989;24:228-33.
23. Gur D, King JL, Rockette HE, et al. Practical issues of experimental ROC analysis. Invest Radiol 1989;25:583-6.
24. Hanley JA. Receiver operating characteristic (ROC) methodology: the state of the art. Crit Rev Diagn Imaging 1989;29:307-55.
25. Metz CE. Some practical issues of experimental design and data analysis in radiographic ROC studies. Invest Radiol 1989;24:235-45.
26. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol 1992;27:723-31.
27. Hecht S. The dark adaptation of the human eye. J Gen Physiol 1920;2: 499-517.
28. Coltman JW. Fluoroscopic image brightening by electronic means. Radiology 1948;51:359-67.
29. Sturm RE, Morgan RH. Screen intensification systems and their limitations. Am J Roentgenol 1949;62:617-34.
30. De Vries H. The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. Physica 1943;10:553-64.
31. Rose A. The sensitivity performance of the human eye on an absolute scale. J Opt Soc Am 1948;38:196-208.
32. Morgan RH. Visual perception in fluoroscopy and radiography. Annual oration in memory of John D. Reeves, Jr, MD, 1924-1964. Radiology 1966;86:403-16.
33. Goodenough DJ, Rossmann K, Lusted LB. Radiographic applications of signal detection theory. Radiology 1972;105:199-200.
34. Rossmann K, Wiley BE. The central problem in the study of radiographic image quality. Radiology 1970;96:113-8.
35. Wagner RF, Brown DG, Pastel MS. Application of information theory to the assessment of computed tomography. Med Phys 1979;6:83-94.
36. Wagner RF, Brown DG. Unified SNR analysis of medical imaging systems. Phys Med Biol 1985;30:489-518.
37. Revesz G, Kundel HL, Graber MA. The influence of structured noise on the detection of radiologic abnormalities. Invest Radiol 1974;9:479-86.
38. Samei E, Flynn MJ, Eyler WR. Detection of subtle lung nodules: relative influence of quantum and anatomic noise on chest radiographs. Radiology 1999;213:727-34.
39. Burgess AE, Wagner RF, Jennings RJ, Barlow HB. Efficiency of human visual signal discrimination. Science 1981;214:93-4.
40. Barrett HH, Aarsvold JN, Barber HB, et al. Applications of statistical decision theory in nuclear medicine. Inf Proc Med Imaging 1988:151-65.
41. Myers KJ. Ideal observer models of visual signal detection. In: Beutel J, Kundel HL, Van Meter RL, eds. Handbook of medical imaging. Bellingham, Wash: SPIE Press; 2000. p 559-92.
42. Eckstein MP, Abbey CK, Bochud FO. A practical guide to model observers for visual detection in synthetic and natural noisy images. In: Beutel J,

Kundel HL, Van Meter RL, eds. Handbook of medical imaging. Bellingham, Wash: SPIE Press; 2000. p 593-628.

43. Tuddenham WJ. Visual search, image organization, and reader error in roentgen diagnosis. Radiology 1962;78:694-704.

44. Tuddenham WJ. The visual physiology of roentgen diagnosis. Am J Roentgenol 1957;78:116-23.

45. Tuddenham WJ, Calvert WP. Visual search patterns in roentgen diagnosis. Radiology 1961;76:255-6.

46. Tuddenham WJ. Perception of the roentgen image. Radiol Clin N Am 1969;7(3).

47. Hebb DO, Favreau O. The mechanism of perception. Radiol Clin North Am 1969;7:381-92.

48. Llewellyn-Thomas E. Search behavior. Radiol Clin N Am 1969;7:403-17.

49. Rossmann K. Image quality. Radiol Clin N Am 1969;7:419-34.

50. Selzer RH. Computer processing of the roentgen image. Radiol Clin N Am 1969;7:461-72.

51. Kundel HL, Revesz G, Stauffer HM. The electro-optical processing of radiographic images. Radiol Clin N Am 1969;7:447-60.

52. Moore R, Ledley RS, Sing HC. Applications of automatic processing methods to the radiologic image. Radiol Clin N Am 1969;7:381-92.

53. Squire LF. Perception related to learning radiology in medical school. Radiol Clin N Am 1969;7:485-98.

54. Tuddenham WJ. Problems of perception in chest roentgenology: facts and fallacies. Radiol Clin N Am 1963;1:277-89.

55. Smith MJ. Error and variation in diagnostic radiology. Springfield (Ill): Thomas; 1967.

56. Berbaum KS, Franken EA Jr, Dorfman DD, et al. Satisfaction of search in diagnostic radiology. Invest Radiol 1990;25:133-40.

57. Berbaum KS, El-Khoury GY, Franken EA Jr. Missed fractures resulting from satisfaction of search effect. Emerg Radiol 1994;1:242-9.

58. Samuel S, Kundel HL, Nodine CF, Toto LC. Mechanism of satisfaction of search: eye position recordings in the reading of chest radiographs. Radiology 1995;194:895-902.

59. Berbaum KS, Franken EA Jr, Dorfman DD, et al. Cause of satisfaction of search effects in contrast studies of the abdomen. Acad Radiol 1996;3:815-26.

60. Kundel HL, Nodine CF, Carmody DP. Visual scanning, pattern recognition, and decision making in pulmonary nodule detection. Invest Radiol 1978;13:175-81.

61. Hu CH, Kundel HL, Nodine CF, Krupinski EA, Toto LC. Searching for bone fractures: a comparison with pulmonary nodule search. Acad Radiol 1994;1:25-32.

62. Krupinski EA. Visual scanning patterns of radiologists searching mammograms. Acad Radiol 1996;3:137-44.

63. Kundel HL, Nodine CF, Krupinski EA. Searching for lung nodules: visual dwell indicates locations of false-positive and false-negative decisions. Invest Radiol 1989;24:472-8.

64. Krupinski EA, Nodine CF. Gaze duration predicts the locations of missed lesions in mammography. In: Gale AG, Astley SM, Dance DR, Cairns AY, eds. Digital mammography. Amsterdam, Netherlands: Elsevier; 1994. p 399-403.

65. Krupinski EA, Nodine CF, Kundel HL. Enhancing recognition of lesions in radiographic images using perceptual feedback. Opt Eng 1998;37:813-8.

66. Tuddenham WJ. Roentgen image perception—a personal survey of the problem. Radiol Clin North Am 1969;7:499-501.

67. Giger ML. Update on the potential role of CAD in radiologic interpretations: are we making progress? Acad Radiol 2005;12:669-70.

68. Kundel H, Krupinski E, Hemminger B. Image perception and workstation design for mammography. Acad Radiol 1999:S215-9.

69. Sowden PT, Davies IRL, Roling P. Perceptual learning of the detection of features in x-ray images: a functional role for improvements in adults' visual sensitivity? J Exper Psych 2000;26:379-90.

70. Lesgold A, Rubinson H, Feltovich P, et al. Expertise in a complex skill: diagnosing x-ray pictures. In: The nature of expertise: Hillsdale, NJ: Erlbaum; 1988.

71. Raufaste E, Verderi-Raufaste D, Eyrolle H. Pertinence generation in radiological diagnosis: spreading activation and the nature of expertise. Cog Sci 1998;22:517-46.

72. Wood BP. Visual expertise. Radiology 1999;211:1-3.

73. Wagner RF, Beiden SV, Campbell G, Metz CE, Sacks WM. Assessment of medical imaging and computer assisted systems. Acad Radiol 2002;9:1264-77.

74. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. Arch Intern Med 1996;156:209-13.