Hist. Sci., xlviii (2010)

PIXELS, PATTERNS AND PROBLEMS OF VISION: THE ADAPTATION OF COMPUTER-AIDED DIAGNOSIS FOR MAMMOGRAPHY IN RADIOLOGICAL PRACTICE IN THE U.S.

Brian Dolan and Allison Tillack
University of California San Francisco

> Some critics have contended that radiologists should now leave all of their technical problems to the engineering staffs of the commercial manufacturers. It seems safe to assume that eventually our radiological tools will reach such a degree of perfection and standardization as to allow us to ignore the details of their construction and operation, devoting ourselves exclusively to the medical phases of our work, but that day is not yet here.
>
> P. C. Hodges, radiologist, in 1945 on the trajectory of X-ray diagnosis[1]

In his book *Why things bite back: Technology and the revenge of unintended consequences*, Edward Tenner says of new technology that "the real benefits usually are not the ones that we expected, and the real perils are not those we feared".[2] This is an apt way to characterize the contingencies, discomforts, and risks that radiologists have encountered over the past century with the introduction of numerous technologies that impacted their professional responsibilities and capabilities of interpreting evidence of the existence (or not) of diseases.

From the introduction of X-rays to clinical practice in 1896, radiologists have explored and adopted the uses of a catalogue of imaging techniques, including the gradual application of ultrasound therapy beginning in the late 1920s, thermal imaging and infrared measurement of metabolic heat in the 1950s, computed axial tomography (CAT, or CT) scanning in the 1960s, and nuclear magnetic resonance imaging (MRI) in the 1970s. Each development involved an array of people with different training and interests before any of the technologies were integrated into the armament of radiological investigations: physicists, engineers, instrument makers and technicians, computer programmers, the military, corporate investors, small animals, physicians, medical researchers in specialist fields, and eventually human patients.

Through examining articles in medical journals, professional association guidelines, medical technology textbooks, and guided by an ongoing oral history and ethnographic study, the present article examines the introduction of a new technology in medical practice by focusing on the anticipated and retrospectively assessed impact it had on enhancing human skills, self-assessed professional competency, and considerations of ethical responsibility in clinical decision making. The technology is an expert system — a combination of programs that provide computer-aided diagnosis (CAD) and artificial intelligence — first designed in the mid-1980s to enable a computer to assist radiologists in the diagnosis of breast cancer. For some specialists, the possible benefits of this were enormous. As was stated in a medical imaging textbook, "computerized analysis offers the exciting option of escaping from

0073-2753/10/4802-0227/$10.00

the anthropocentric description of images, and go beyond the limitations of the human visual and cognitive system".[3] As we will see, however, others were worried about the implications of 'escaping' anthropocentric analysis and extending responsibility for clinical decision making to newly defined technological 'experts' — whether software designers or artificial intelligence systems themselves.

Our case study of the development and impact of this medical technology builds on the historiography of visualization in science and the history of machine assisted vision. Our conclusions lead to more recent insights about how technologies that are designed as adjuncts to human skill affect practice protocols and judgements about disease ontology and ideas of tissue "normality" — which for breast tissue appears to have a unique history of problematic determinations.[4] The background literature is extensive, and many different theoretical approaches to the topic prove to be relevant to this case. The historical interconnections between the visual arts and biomedical sciences have been elaborated on in a number of essays in the valuable edited volume *Picturing science, producing art* by Caroline A. Jones and Peter Galison, Martin Kemp's collection of essays on how scientists see the world in *Seen / unseen*, and Barbara Stafford's recent work which explores how images and patterns can themselves be used to illuminate the neural work of human consciousness.[5]

The mediation of visualization and cognition through instrumentation in science has been widely investigated over the past few decades. Studies of particular use to us include works in the history of microscopy by Jutta Schickore, Alberto Cambrosio and Peter Keating's account of flow cytometry and the visualization of lymphocytes, and Brian Dolan's work on the problem of reproducing colour in natural history manuals for specimen collection.[6] Finally, the ways in which image production technologies (re)define disease, define the ontological status of invisible entities, or otherwise make epistemological or legal meaning of images are discussed insightfully by Ian Hacking in *Representing and intervening*, Nicholas Rasmussen's work on electron microscopy, and Lorraine Daston and Peter Galison's recent opus on *Objectivity.*[7] Of particular note to our study of the digitization of X-ray mammograms and how this fits into the historical evolution of the interpretation of X-rays is Andy Warwick's recent article on "X-rays as evidence" and Bernike Pasveer's work on the sociology of X-ray interpretation.[8]

Any study that investigates the development and uses of technologies of visual production can benefit from this broad historiographic framework, and we certainly have. But while we see our study as a contribution to the way we understand the relationship between clinical images and diagnostic decision making in the production of medical knowledge, the value of our case study lies in its depiction of how clinical practice finds ways of adapting to the uses of emerging technologies upon the moment they are "black-boxed". Our study grapples not only with the implications of the computerized production of images and the challenges of interpreting digital, rather than conventional analog, X-ray images of breast tissue. It also raises questions about the social and ethical implications that emerge with the addition of artificial intelligence that is used by computers to "interpret" *their own* digital productions and advise clinicians on what they should be doing. Thus we attempt to move beyond familiar accounts of human deskilling and discussions of the "technological invention of disease" to reassess how human cognition and decision making is itself challenged in the face of a technology ostensibly modelled on what is recognized as an imperfect human skill: interpreting X-rays.[9] While not a history *per se* of breast cancer, our study also demonstrates how very particular the social, technological and professional context was that made computer aided diagnosis possible — if also controversial — as a clinical tool for breast cancer screening in the 1980s and 1990s. We wish here to emphasize that while we see CAD in mammography as both a product of a social and professional concern for national breast cancer screening and a technology that problematizes the integrity of that protocol, our case study is less about attitudes toward breast cancer screening than about how credibility for a diagnostic technology used in that arena was negotiated.[10]

SHADES OF GREY

In the 1980s and 1990s radiologists were grappling with problems of interpreting conventional X-ray mammography and confronting risks of inaccurate diagnoses. The skill at the heart of radiologic practice is interpreting images. Depending on the technology used to produce them — whether X-rays, ultrasound, MRI, etc. — this involves differentiating potentially pathological from normal tissue based on variations in shades of grey on film, sound waves' echoes on a screen, or coloured pixels on a monitor. This information is then correlated with other data. The radiologist's interpretation of the image will predominately determine whether or not the patient will be subjected to further tests, and possibly to surgery.

The history of mammography, its techniques of film-screen processing and the attention paid to the procedures by the American Cancer Society, the National Cancer Institute and the American College of Radiology, especially after the Breast Cancer Detection Demonstration Project in the 1970s, is well known — Baron Lerner's 2001 book *The breast cancer wars* provides an excellent overview as does the more recent account by Robert Aronowitz *Unnatural history: Breast cancer and American Society*.[11] There are just a couple of points necessary to mention here regarding the recommendations that women regularly have biennial base-line mammograms after age thirty-five and annually after age fifty.[12] They point to a context that affected radiologists' views about the applicability of enrolling new computer technology for assistance in their work.

First, radiologists expressed concern about potential problems of an increased workload generated by a national screening program. With an estimated 47 million women of screening age in the United States in 1990, it was stated that "every radiologist, regardless of expertise, would interpret 2,350 mammograms per year, or 9 per day".[13] Studies in the development of PAP smears by Vicky Singleton and Mike Michael as well as Adele Clarke and Monica Casper have shown how complicated screening programs and classification practices become in such instances.[14] Screening for breast cancer was no exception. But unlike other screening tests for cancer, such as those for cervical (Pap), colon (stool guaiac) or prostate (prostate-specific antigen), where the prior probability of disease was pre-screened by clinicians, breast cancer screening was primarily an imaging test, requiring the interpreter to have "a talent for perception of abnormalities" because the procedure carried "a degree of difficulty that is underestimated by both our critics and our colleagues".[15] It was a technique designed to detect "invisible" tumours, and as is made clear by a number of authors and public debate, it remains controversial to this day.[16] And similar to the work on investigating other forms of diagnosis, here we also see how the integration of new technologies destabilizes the professional boundaries of those involved with its development and uses. Implicated in the reconfiguration of such networks are evaluations of human skill and the cognitive ability of artificial systems — how "intuition" versus "calculation" is weighted in medical decision-making.[17] According to a radiologist writing in the *New England journal of medicine* in 1986, because of the amount of work necessary to properly read mammograms, "it will be necessary, in my opinion, to use non-physician radiology assistants to interpret uncomplicated examinations. The potential political and legal ramifications of doing this are obvious".[18] While such ramifications are indeed important, and we will briefly return to the political and legal ones toward the conclusion, there are also practical and ethical ramifications of dividing labour in this way, not least since what constituted an "uncomplicated examination" was of course not self-explanatory. In the 1980s and 1990s computerized systems were introduced as a possible way of compensating for the limitations of human interpretive skills. Such systems purportedly worked by differentiating to degrees imperceptible to human vision variations in image patterns by assigning values to shaded pixels in digitized X-ray film. In addition, artificial neural networks (ANN) were programmed to "learn" by example, and eventually recommend a course of action to the radiologist based on the computer's interpretation of its own digitized image. This article, therefore, further explores the challenges involved with making medical decisions that rely upon a computer-generated image that extrapolates data from segments using instructions that necessarily conform to a previously defined set of possibilities of what the image could look like.[19] The intention behind computer-aided diagnosis (CAD) systems was to remove much of the complexity of interpretation, and the time spent reviewing thousands of images, by transferring human skill to another agency. Precisely because diagnosis concentrated on pattern recognition in images, it was suggested that a combination of new digital technologies could be implemented to allow a computer to perform the job of an expert radiologist. The introduction of a new "expert system" in medicine was hailed as the beginnings of a possible revolution in computer aided diagnosis. A number of historical studies on standardization and rationalization of medical practice inform our understanding of different ways that data, or clinical evidence, is made to look unbiased and objective.[20] For instance, in his book *Rationalizing medical work: Decision-support techniques and medical practices*, the sociologist Marc Berg proposes that interest in the development of "expert systems" in medicine revolves around the conceptualization of medical practice as a scientific, hypothesis-driven endeavour.[21] Berg notes that the locus of problems with medical practice shifts over time, from external to the medical field in the post-war period to within medical practice in the 1960s and '70s to *within the physician* in the 1980s and beyond. Along with the rise of evidence-based medicine and the scientific method as the model for medical problem solving was the "cognitive revolution" in psychology that argued that human minds function as information-processing systems.[22]

These changes triggered a profound shift in how medical practice is conceptualized. Berg states that "medical work is not seen as a social activity, but as an individual, cognitive process" in which physicians' reasoning is reflective of the scientific method.[23] However, the human brain is viewed as fundamentally flawed and deficient in this new paradigm: the wasting of medical resources and poor quality of care are attributed to the mental incapacities and suboptimal decision-making capabilities of individual physicians.[24] Attributes highly valued in the scientific method, such as memory, objectivity, and probability, are believed to be performed better by machines than humans, who are portrayed as biased, easily distracted, and reliant on personal experience rather than scientific fact.

CAD programs designed to "read" mammographic films were not only developed in the context of trends toward viewing physicians as information-processors guided by the scientific method. Additionally, medical images such as X-ray mammography are situated in an environment where machine vision has gained moral authority and attributed a "mechanical objectivity" that is often deemed superior to human vision. Datson and Galison's research on the evolution of the scientific atlas, like the research by Galison, explores how the concept of objectivity is intimately linked to medical imaging.[25] They argue that objectivity is a historical concept, intertwined with depictions of visual images and changing values concerning the role of the scientist. The notion of objectivity became a "shifting border between judgement and mechanization, between the possibility (or necessity) of human intervention and the routinized, automatic functioning of the technology".[26] The knowledge of a scientist and the ability to form judgements based on that knowledge made up the core of scientists' claims to authority and authorship.[27] However, judgement is ultimately "an act of perception and cognition ... associated with a picture of reading that is both anti-algorithmic and anti-mechanistic".[28]

Considering the trend described by Berg towards viewing modern medicine as an "objective" science based on statistics and epidemiological evidence rather than the subjective judgements of physicians,[29] it is not surprising that radiologists themselves became viewed as the "design fault" in medical imaging. It is also not surprising that many radiologists were slow to embrace technologies such as CAD; by transferring knowledge and judgement to a computer, the authority and intellectual territory of radiologists were threatened.

CAD was designed to confront challenges of human "subjectivity" (in this case variation in human visual ability) and the infinite variation of organic pathology. By examining the evolution of mammographic CAD, we see how the transformation from analogical to digital radiologic imaging replaced organic concepts of disease with algorithmic calculations of deviance from mathematically constructed models of normality. Pathology is re-classified by relying less on "intuition" and judgement and more on algorithmic calculations of abnormal variation, which in turn creates a black box dictator that brings into question the responsibilities of medical decision-making. In turn, the evolution of computerized visual modelling of normality and pathology impacts the professional practices and professional identity of those involved with new ways of perceiving. A new kind of expert is born.

REDEFINING HUMAN VISION AND RADIOLOGICAL EXPERTISE

The increasing dependence of biomedicine on medical imaging and the implementation of population-based screening programs using X-ray images have ensured the continuing interest of researchers in the development of radiologic expertise and in trying to answer the question "What makes a good radiologist?".[30]

The field of psychology has been interested in the development of expertise since the 1960s. Viewing expertise largely as a function of memory, the chessmaster deGroot employed what he termed "protocol analysis" to examine how highly skilled chess players functioned in comparison with novices.[31] He asked players to verbalize their thoughts while playing, and then carefully analysed these thought processes with the hope of distinguishing differences in thought protocol between expert and less-skilled players. While deGroot's work did not reveal substantial differences in the verbalized thought processes of players at different skill levels, Chase and Simon built on his analysis by studying the ability of expert chess players to store and retrieve information.[32] They found that experts were better able to store and retrieve "chunks of information" which were organized into meaningful mental representations or schema. Critically, they found that these highly skilled players were not consciously aware of the process of constructing these schemas or using them as a tool to organize and retrieve information. In their extensive treatise on expertise, Ericsson and Smith suggest that experts are not just storage vessels for a large amount of complex data.[33] Instead, Ericsson suggests, experts are able to select relevant information and "encode it in special representations in working memory that allow planning, evaluation, and reasoning about alternative courses of action".[34]

Interestingly, further studies of a wide variety of experts seemed to indicate that expertise is domain specific.[35] For example, while expert radiologists are able to identify pathology on a chest X-ray far more accurately than a medical student, the experts are no better at general tasks requiring visual searches, like finding Waldo in a children's book.[36]

Initial studies of expertise focused on such activities as chess, interest in expertise and skill quickly expanded to other fields.[37] The development of X-ray images was closely followed by the observation that the interpretation of those images can vary widely. The question of how radiologists develop expertise in image analysis and why variation in image interpretation exists has yielded many interesting studies on how expert radiologists "see".

The psychologist Lesgold and his colleagues use Chase and Simons' study as the foundation for their examination of the work of expert radiologists.[38] They argue that radiology is an unusually complex and difficult skill that requires the integration of perception and detailed visual analysis with an in-depth knowledge of anatomy, pathology, physiology, and physics. Like Chase and Simon's chess masters, Lesgold proposes that when faced with a problem (i.e. an image that must be interpreted) expert radiologists utilize organizational schemas which guide further evaluation of the problem. "Perception", Lesgold *et al.* state, "is driven by mental representations".[39] By asking radiologists with different levels of experience to view chest X-rays and verbalize their thought processes while analysing the images, Lesgold proposes that experts are able to identify the relevant schema during the first few seconds of viewing an image. This schema is then used to guide where the radiologist looks and enables him or her to form a basic representation which can be tested with the data that are available. Lesgold *et al.* also found that expert radiologists were more efficient and "flexibly opportunistic" than residents with only a few years of training.[40] Critically, they state that skilled radiologists are able to adjust notions of normal anatomy to fit specific images, thus allowing the identification of pathology. Novices, however, rely on fixed "localization cues" that may misrepresent normal anatomic variation as pathological. Unlike novices, experts are also able to incorporate new data such as test results or patient history and use these data to refine or adjust the appropriate schema.

According to Lesgold, expert radiologists are able to combine flexible mental representations or schema of data with fine-tuned visual acuity that enables detailed discrimination of image features and the ability to consider multiple interpretations of a single image. "First, a perceptual decision is made, the outcome of which is a differential diagnosis set with associated probabilities. Then, cognitive decision-making apparatus is used to resolve ambiguity, either by searching for perceptual features initially missed ... or by taking account of other data sources such as history and tests."[41] According to Lesgold and colleagues, the process of acquiring radiological expertise involves movement from a superficial, probabilistic approach to "deep reasoning" through the continuous refining of schemata and development of an automatic capability for pattern recognition.[42] Interestingly, in their study they found that radiologic performance is not simply a function of experience. Advanced radiology residents were less accurate in their diagnoses than novice residents or experts; Lesgold *et al.* posit that this seeming "regression" in skills in fact represents a shift from relying on superficial associations to the development of an advanced understanding of images based on pattern recognition.[43]

HUMAN VISION AND ITS LIMITATIONS

According to a 1986 survey by the American College of Radiology, the majority of radiologists were using combination "film-screen" mammography, a technique developed in the early 1970s which replaced non-screen (direct) film mammography and was an alternative to xeromammography. Favoured for its relatively low radiation doses, its rapid processing, shorter exposures, and sharper images, film-screen mammography involves placing a sheet of film on an intensifying screen (hence "film-screen") under the breast, while a plastic compression plate (with a grid system of holes in it, visible on the mammogram, to guide needle placement if biopsy is necessary) is lowered from above which prevents breast movement and improves image quality by separating tissue and reducing "scattered radiation". Decreased breast thickness also reduces radiation doses, since an X-ray beam is directed through the breast, leaving an image on the film. Tissues within the breast, including any microcalcifications (a small calcium deposit accumulated in breast tissue), dense masses, cysts, and carcinomas absorb the X-ray photons and appear on the film as bright spots.

In younger women, particularly under age 40 who have not given birth, a "normal" breast is "radiopaque" — its mammographic image appears predominately white since the breasts are composed of dense fibroglandular tissue. With increasing age and childbearing, the breast tissue is replaced with "radiolucent" fat, which does not absorb the X-rays and produces a more translucent image. Dense tissue, as well as surgical scars, fibrocystic changes, and calcific-like deposits on the skin from tattoos, deodorants, or ointments are among the common factors that limit the accuracy of mammographic examination since any cancer, which appears as a white mass on the X-ray film, is cloaked by the white mass of other dense tissue or matter; they are therefore referred to as "occult breast tumours".[44]

Other, external, conditions also affect the quality of the mammogram. As a professor of breast imaging at UCLA explained, "radiographic contrast" — the differences in optical clarity between different areas of the film — "depends on subject contrast (radiation quality, kVp), film contrast (a property of the X-ray film), film processing (darkroom conditions, development temperature, chemicals and time), and scatter reduction ([breast] compression, grids)".[45] The physical construction of the mammographic station — the imaging apparatus — and the exacting nature of the film development process often create noise, "artifacts", or phantom images on the resulting mammogram. Thus, the combination of breast physiology, clinical, and laboratory conditions may affect the quality of the mammogram that a radiologist examines as a crucial diagnostic test for breast cancer.

In 1986, the American Cancer Society suggested that some standard be set for technical specifications of the equipment, type of film used, and ancillary devices, which at the time varied according to the preference of the radiologist. This was an important point, explained Philip Strax, medical director of the Guttman Breast Diagnostic Institute in New York, since "cancer may be obvious with one set of technical circumstances and not with another".[46] When the American College of Radiology established its Mammography Accreditation Program (MAP) the next year, 1987 (a peer-review program meant to be the standard by which mammography facilities were deemed acceptable), it further spelled out the conditions under which radiologists should view the mammograms. Their recommendations included adequate viewbox luminance, low ambient room light to reduce reflection off the film, and masking the view box to stop back-light from flooding the eye.[47] These measures were intended to improve the conditions that affected radiologists' interpretive skills and to persuade physicians of the reliability of mammography.

"The correct interpretation of a properly performed mammogram is the ultimate contribution of the radiologist", stated Strax. However, it was evident that not enough attention was being given to the enhancement of visual skills. "Many technologists develop considerable expertise in interpretation", noted Strax, referring to non-medically trained radiology assistants. "This expertise should be used by the radiologist as a back-up. There undoubtedly have been many instances when the sharp eyes of a technologist have saved the radiologist from possible oversight."[48] It was clear that the repertoire of medical skills required of a radiologist now included learning to recognize disease through rote practice more in step with what they perceived as the "routines" of their technicians, a bias that has been investigated by historical sociologists where discrepancies have been found between the status accorded to technicians and the level of deep, situated knowledge they display in everyday work.[49]

In 1990, only an estimated 500 out of 8,000 mammography facilities across America had been given accreditation by the American College of Radiology, but the passage of the Mammography Quality Standards Act by Congress in 1992, under the enforcement of the FDA, helped further develop standardized mammographic protocols.[50] However, as much as this improved the uniformity of diagnostic practices, further studies were bringing forth unsettling findings regarding non-uniformity of interpreting mammograms.

A 1990 study, for instance, recognized that "Many missed radiologic diagnoses can be attributed to human factors such as subjective or varying decision criteria, distraction by other image features, or simple oversight. Studies suggest that these errors may be inevitable with human observers, and that they are not strongly related to experience".[51] Time, it was suggested, was the essence of the problem. The authors stated that "the time needed for screening four-view mammographic study for significant pathology to be as short as 45 seconds." In a clinical setting where a high volume of films was reviewed, such as a screening program, forty-five seconds is hardly a meditation (by comparison, in 1989, a cytologist wrote that "a careful reading of a Papanicolaou smear requires at least five minutes per slide … and a difficult case sometimes requires considerably more time").[52] Furthermore, the investigators did not think that double-readings — having two radiologists routinely examine the same film — were particularly effective either, especially when cost and time-efficiency of work in a busy clinic were considered. "It is in such conditions", they said, "that observer oversight might be expected to play a significant role in missed diagnosis".[53]

A subsequent study conducted the following year, 1991, probed deeper into the extent to which radiologists differed in their interpretations of patients' mammograms.[54] Ten radiologists who "routinely read mammograms" separately examined 150 images "considered to be of good technical quality by 1987 standards" — the year that the images were obtained, and which like the previous study were produced with a standard film-screen technique. The interpreters were asked to select from three diagnostic categories: "normal", "abnormal, probably benign", and "abnormal, suggestive of cancer". The results were alarming. The median weighted percentage of agreement was 78%, yielding "moderate" diagnostic consistency. But in 25% of the comparisons for the group of women as a whole, there was a substantial disagreement in patient management recommendations, in which one radiologist recommended routine follow-up while another recommended a biopsy for the same patient. In 9% of the cases, radiologists agreed in their recommendation of a biopsy, but disagreed on whether it should be for the left or the right breast. The report's conclusion seems obvious: "Efforts to improve accuracy and reduce variability in interpretation may increase the effectiveness of mammography in detecting early breast cancers."[55]

It was suggested as well that performance levels could be improved if more time were spent in training. This is supported by the results of a 1992 study on "improvement in mammography interpretation skills" (in a community radiology practice) which showed a marked improvement in a dozen radiologists' data evaluation and cancer detection skills after they attended seventeen dedicated mammography courses over a two-year period (1987–89).[56] This led the investigators to conclude that "the major factor" for this was their training in additional image evaluation (chance factors were discounted on the grounds that there was no difference in the quality of images). As plausible as the suggestion was that continued re-training could enhance interpretive skills, once again issues of the cost and time involved led to doubts that the scheme would prove effective. Indeed, it is suggestive that the quality assurance guidelines never prescribed repeated training courses.

It was at this point, however, in the midst of a major political and medical campaign spearheaded by the two largest cancer awareness groups and amidst growing recognition about the demands and difficulties of interpreting conventional mammographic images, that computer-aided diagnosis was introduced as a solution to radiologists' and patients' problems.

The promises of the new technologies included faster screening, higher resolution images, a release of the demands on the time of radiologists, and higher levels of diagnostic accuracy. However, the proof of the success of an automated, computerized approach to breast cancer screening was less readily available. Furthermore, it occurred in the midst of challenges to the profession of radiology relating to the assessment of their skills and their decision-making capabilities. The concept of relying on a machine for what could be life-and-death decisions therefore raised even more challenging questions about the very nature of medical practice and the conditions that guide decision-making processes.

COMPUTER VISION AND ITS LIMITATIONS

Throughout the early 1980s, the literature on computer-aided diagnosis was scarce, but with new developments in computer technology, especially laser scanning and printing, mammography entered a phase of rapidly "going digital". This involved scanning conventional film-screen mammograms, using a computer program to enhance clusters of pixels with shapes and colours of interest (colours being shades of grey), and print out a high resolution image for the radiologist to review. In 1987, a group from the Laboratories for Radiologic Image Research, at the University of Chicago, published the first of a number of articles that discussed CAD with reference to the potential uses of new computer technology. As they stated, "The efficiency and effectiveness of the screening process may be greatly increased if an automated computer system can be successfully employed for the detection of microcalcifications".[57]

The research group's computer-aided approach involved multiple stages of image acquisition and enhancement. In brief, the first step was to obtain a digitalized screen-film mammogram with a Fuji drum scanner which produced an image with 1024 grey levels with the size of the region that contained the breast image being approximately 1000 × 1800 pixels. Once the digitalized image was obtained, it was then re-processed with two filters. The first filter produced an image with enhanced characteristics: enhanced pixels that matched the pre-programmed size and contrast variations of "a typical breast microcalcification". But, as the team noted, "since the size and shape of microcalcifications vary, it is not possible to design filters that exactly match each different microcalcification".[58] Thus the "match" between a cluster of pixels with a certain size, shape and grey-scale value and a microcalcification necessarily relied upon an approximation of what a "typical" microcalcification looked like in the mammograms the radiologists had previously studied.

The second filter did the opposite to the first. With a "signal suppression" filter, the pixels that had values representing all non-microcalcification characteristics were kept, while everything else — which corresponded to what might be microcalcifications — was eliminated from the image. These two filtered images were then "subtracted" from one another, producing a "difference image" — a new digitalized image that was derived from the original mammogram. At this point, the digital image is only as useful as the programming that guided the double-filtering process. Because the "difference image" is enhanced and altered according to the way the computer is programmed to "see" (or ignore) microcalcifications, any other characteristics of the breast that existed can no longer be considered part of the image. This, of course, is the point of computer-aided diagnosis centred on identifying microcalcifications: the program follows instructions rigidly, and is allegedly never "distracted" from its field of vision. Indeed, the next stage in digital enhancement is to produce a "threshold image", in which groups of two or three pixels with values corresponding to what approximates microcalcifications (which "generally" are less than half a millimetre in length) are superimposed on "an absolutely uniform background".[59]

What remained to be tested, however, was whether the programming (which assigned values to the shape, colour, and size of pixel groups) provided a reliable and accurate guide to identifying microcalcifications. To determine the true-positive and false-positive results, the research team used a computer program that generated simulated microcalcifications and placed them on the images for the computer algorithm to detect. The results were as high as 90% true-positive detection (whereas unaided radiologists fell between 70% and 90%). But that was a simulation. Further tests needed to be done to determine, for instance, how to prevent the computer from misidentifying other breast structures such as fibrous strands, ducts or skin folds which have similar appearance (and therefore might be assigned the same values) to microcalcifications. This would involve substantial amounts of programming, and at the time, the "computational requirements of digital mammography and computer analysis of mammographic images limited the practical application of their techniques".[60] However, it did not take long for fresh enthusiasm to emerge, which occurred when a new technique was developed in the field of artificial intelligence in medicine (AIM).

ADDING INTELLIGENCE

While new to the field of mammography, computer systems designed as complex "problem solvers", or "expert systems", had been introduced to aid medical practice over the previous twenty years. In 1980 it was estimated that there were some 2,000 articles describing various automated medical decision-making and consultation systems.[61] Systems such as INTERNIST, which aided diagnosis in internal medicine, MYCIN, which was developed to provide consultative advice on diagnosis and treatment of infectious diseases, and others, while mostly experimental rather than clinically applied, were nonetheless recognized as an important mechanism for improving access to high-quality health care. At the same time, computer scientists and medical anthropologists who studied the medical profession to determine how best to emulate clinical cognition and "formalize" human expert knowledge (acting as "investigative reporters" who study what physicians know and how they solve problems), described the ability of the computer to store extremely large amounts of data, to enumerate many possibilities for classifying data, and to perform complex logical operations thereby underscoring its potential value in medical problem-solving processes. Such systems, it was proposed, would provide a standardized logic to organizing the explosion of information about pathophysiology that might send clinicians into "future shock" — frozen with the proliferation of diagnostic and therapeutic technologies.[62]

It was in this context — public attention to promises of improved health care, experiments with clinical algorithms (protocols) that were gaining the attention of physicians and patients,[63] and practical improvements in computer functionality (especially with the development of microprocessors facilitating lower costs and mobile technology) — that a new expert system for computer-aided diagnosis in mammography was born.

Three years after the 1987 publication describing the uses of computer-aided diagnosis in mammography (described above), the same research group at the University of Chicago published the results of a number of tests assessing the potential usefulness of artificial neural networks (ANN) to assist diagnosis — aiding observation in the areas of chest radiology and digital mammography. A neural network was, the research group explained, "a computational model based on the brain; it is a powerful tool for pattern recognition".[64] In much the same way that doctors were understood to make decisions — by weighing evidence presented to them, drawing on past experience with similar cases, then making a diagnostic prediction — they explained that a neural network could be programmed to evaluate evidence and even learn from its own experience.

ANN consists of a set of processing units (called nodes) which are interconnected via paths that allow inputted data of different "weights" to travel through the network in parallel as well as serially, performing a non-linear calculation (analogous to neurons and synaptic connections in the human nervous system). Each incoming signal to a node — each piece of information from a dataset inputted to the neural network — is given a numeric value (a "weight") assigned by certain highly skilled medical professionals (who become the "gold standard"). The neural network is initially programmed by inputting a certain number (the higher the better) of examples whereby each piece of clinical information that went into making a correct diagnosis (patient's age, sex, symptoms, and array of radiographic signs) is weighted and the neural paths accordingly programmed to yield a correct diagnosis. Then, each time the computer is fed a dataset, it (like a young medical student) can weigh the evidence according to what it has learned from the previously inputted examples and provide a diagnosis. If it encounters information that it has not been programmed to weigh, or the weight of the sum of evidence does not have a predetermined output path, and it provides an incorrect diagnosis, "these new data can be incorporated into the data base along with the correct diagnosis so that very similar cases, which may be encountered subsequently, will be correctly identified".[65] In this way, the neural network is said to "learn" from its own mistakes.

In a number of studies on the computerized detection of clustered microcalcifications in digital mammograms,[66] the team from the University of Chicago had an expert radiologist locate "true" microcalcifications in digitized mammograms. Forty-three radiographic features defining calcifications were selected as weighted inputs to the neural network. These features ranged from the shape, size, and pattern of breast masses to the number, uniformity, and distribution of calcifications. Then, 133 "textbook cases" were selected from a published mammography atlas as a training data base, whereupon an experienced mammographer assigned a value to each of the forty-three features and programmed the correct diagnosis (the "truth"). To test the system, sixty separate clinical cases showing abnormalities that were independently proved to be either a mass, a cluster of microcalcifications, or another abnormality, were fed into the computer system, and its performance was compared to the interpretation of attending radiologists. The neural network performed with higher sensitivity (probability of diagnosing a malignant lesion), higher specificity (probability of correctly diagnosing a benign lesion), and higher positive predictive value than the average performance of the radiologists. "Therefore", the research team concluded, "the neural network, working with features extracted by an experienced mammographer, appeared to be able to recommend an appropriate course of action better than the average radiologist, the average resident, or the experienced mammographer himself".[67]

The results were encouraging to advocates of CAD. It promised to give legitimacy to the process of identifying diseases and acting on their treatment. The message resonating in dozens of articles published between 1990 and 1995 was that ANN was simply far smarter than any human.

Although the use of ANN as an automated classifier was celebrated by some, there remained drawbacks and untested parameters, most significantly being the facts, first, that the values of the extracted features used for training were not exhaustive, and second, that they relied on the subjective interpretation of the inputting radiologist. In other words, there was no comprehensive or standardized way to classify the digital representation of disease. This issue had emerged when mammograms were first digitized for computer analysis, where the brightness of each pixel was assigned a value (the "grey level") which was recognized and analysed by the computer for the "enhancement of *meaningful structure*, the quantitative description of image characteristics and features, or the detection of *abnormalities*".[68] The introduction of apparently self-learning artificial neural networks did little to draw agreement about what "meaningful structures" and "abnormalities" look like. How does one test the system if the gold standard is still the human interpreter whose own limitations are precisely what the computer is to transcend?

The philosopher Hubert Dreyfus and the engineer Stuart Dreyfus, who have written extensively on skill acquisition and computer-based expert systems, argue that artificial neural networks are fundamentally unable to capture and reproduce expert behaviour, especially in the field of medicine.[69] These systems rely on a combination of "book knowledge" (those facts that can be found in textbooks and journals and entered into the computer) and "heuristic knowledge" (which comprises "working rules of thumb"[70] or the knowledge gained by practice and experience used by radiologists to make decisions). Dreyfus and Dreyfus state that while novices depend on rules and facts to perform a task, the expert's vast experience allows him or her to intuitively act without rules. By asking an expert radiologist to verbalize the heuristics or rules that he or she uses to interpret an image, the expert is being forced to remember rules he or she no longer uses and therefore regresses back to the level of a beginner. Thus, according to these authors, a system relying on this kind of design can achieve competency through the combination of rules and facts, but can never replace the human expert.[71]

Interestingly, similar conclusions have been suggested within the field of radiology concerning the use of CAD in mammography. In her editorial in *Academic radiology* on the future of CAD, Krupinski writes that "there may be features in the image that the radiologist is not necessarily cognizant of, features that the human visual system is attuned to and uses during image search and interpretation".[72] In addition to recognizing this "tacit knowledge", Krupinski further acknowledges that machine vision cannot replace human vision and may in fact "see" in a fundamentally different way. "There are certainly lesions that the computer can detect and the human observer cannot ... there are also lesions that the human observer can detect and the computer cannot".[73] Krupinski states that "the combination of CAD plus radiologist will yield better performance than either one alone and that the radiologist will be able to rely on his or her own perceptual and cognitive systems, training, and experience to decide the validity and importance of each mark that CAD provides".[74] As Berg predicted, instead of a program that can detect mammographic abnormalities more accurately and efficiently than a radiologist, the introduction of CAD into mammography has reconfigured both the tool and the user. In fact, in recent publications the abbreviation CAD no longer stands for "computer aided diagnosis" but rather "computer aided detection", underscoring a shift in perception about what these programs are capable of achieving.

Nevertheless, some researchers have attempted to derive classification systems informing risk assessment based on pixel analysis, none of which is universally accepted. However, a relatively simple method was developed in 1994 for a computer to sort images automatically into two categories: those which are "easy" and those which are "difficult" to interpret, referring both to the interpretive capabilities of humans and the machines they program.[75] The usefulness of sorting images in this way was meant to quickly facilitate the decision of which cases should be referred to "the most experienced readers" and those which an assistant radiologist could interpret. While it was recognized that this "could permit the better use of the time and skills of expert radiologists", what made the category of "difficult" images additionally useful was that it would provide research materials to continually test the capabilities of the artificial neural networks.[76]

This worked in the following way. The computer-designated "difficult" images were subsequently assigned the truth (positive or negative for cancer) and the data re-entered in the program to create a new dataset for refined future performance. This was a step along the way to creating what a research group in 1999 considered a diagnostic aid that went beyond human perceptual features in the first study to evaluate the clinical potential of computer classification based on computer extracted features completely independent from radiologists' interpretation of mammograms.[77] Such work, which other radiologists have described in terms that make it clear that computers will always supplement, and not replace, human experts, nevertheless suggests a possibility that computer-aided diagnosis could assume a status whereby their operations will be taken for granted by virtue of the fact that they can assimilate and calculate more quickly than any human brain, and work "independently" and without fatigue.

Yet, for all the hype, the acceptance of the role of intelligent computer vision in medical practice relied on "humanizing" the machine. The appeal of neural networks seemed to lie in the way that their construction and operation mystically approached an understanding of the workings of the human mind. Programmers avoided portraying the computer's functions as merely an elaborate algorithm, which might have lingering associations with crude protocols developed in the 1960s to help paramedics make emergency decisions. As was expressed in a *Lancet* editorial discussing the potential for such expert systems, ANN "suggests kinship with the deepest secrets of human thought, … [having attributes] of 'hidden neurons', 'perceptions', and 'synaptic weights'".[78]

Portraying the machine as super-human rendered the appearance that there was something more meaningful in its construction than a mere set of rules which were rigidly followed; indeed, rules of behaviour that could not be broken. A rule-governed computer system — incapable of "internalizing" inexpressible skills attributed to human experts — could only mimic a novice's rule-bound, "behaviour-specific" actions.[79] As such, it would be perceived to have practical value no greater than the "sharp eyes of a technologist" who, as we recall Philip Strax saying, "have saved the radiologist from possible oversight".[80] While useful as an extra set of eyes, such limited expectations of the capabilities of ANN or computer-aided diagnosis fell well short of the idealized view of the systems' designers.

However, those who did not accept that computers were superior to humans — just as technologists were not considered to be *better* than expert radiologists at decision making (but were good at catching the occasional error) — expressed another reason why caution should be taken in the use of such systems which pointed to the ethics of deferring diagnostic logic or clinical cognition to another agency. "Merely feeding clinical data into a computer and reading the result", explained a clinician writing in the *Lancet*, "irrespective of the method used to derive it, could undermine the clinician's ability to take personal responsibility for clinical decisions".[81] In particular, according to another contributor in the same journal, "in negligence cases courts may consider black-box systems as products not services, forcing developers to take on strict liability because they interfere with the ability of a professional to act as a 'learned intermediary'".[82] Thus, however purportedly intelligent such "black boxes" were, they faced difficulty gaining acceptance since the designers conceived of the benefits of the support tools according to their own idealized view of how medical decision making should be done, rather than being based on the needs of the clinicians and their patients.

Rather than embraced as a welcome aid to their practice, some radiologists considered such systems an "interference" and more burdensome on their time. Even the foremost advocates of CAD acknowledged that "techniques for the computer detection of mammographic abnormalities vary markedly in their structure and execution … [and] require that a number of empirical decisions be made regarding parameters that occur during the execution of the program".[83] And finally, once all the work had been done in the attempt to standardize and calibrate the machines to enhance their performance, the concern then becomes one of how much influence the "computer vision" will have on human vision. While designers of intelligent systems assured the radiology community that "the radiologist using the workstation will always make the final decision", new studies were beginning to find that computer-aided "prompting may affect both the performance and the visual search behaviour of radiologists interpreting mammograms".[84] Once the computer suggests a "region of interest" to examine, a behavioural pattern emerged whereby human interpreters merely followed the machine's instructions for where to look, spending little time elsewhere on the film, leading to potential oversight of other features that would be searched for in conventional mammography. It is clear that in 1995, fifty years after the prediction that "our radiological tools will reach such a degree of perfection and standardization as to allow us to ignore the details of their construction and operation", that time had not yet arrived.[85]

CONCLUSION

Following Berg's analysis, it is no surprise that creating a computer "expert" that would be free from the "constraints" of the human mind was proposed as a solution to the mammography crisis. A machine could work without distraction or need for rest, and would not be hampered by flaws in human thought or subject to "the inherent boundaries ... in human perceptual processes".[86] Yet, as Berg points out, getting a tool such as an expert system that detects mammographic irregularities to "work" is a complex process: "tools do not simply slip into their predestined space within a practice. Rather, getting a decision-support tool to work and constructing a niche for it in a local medical practice involves continuous negotiations with all the elements that constitute the practice."[87] Both tools and their users are transformed during this process, often yielding unanticipated consequences. The integration of new technologies destabilizes the professional boundaries of those involved with its development and uses. Implicated in the reconfiguration of such networks are evaluations of human skill and the cognitive ability of artificial systems. While analyses tend to focus on what is lost when a human skill is transferred to a machine, it is also critical to consider what is inadvertently *gained* once the machine is built. As we see, the identification of a "region of interest" by CAD programs can alter a mammographer's behaviour; in this case computer vision directs human vision. The question of who has control of the encounter between radiologist, image, and computer remains unanswered; the radiologist is influenced by results generated by CAD, but at the same time the radiologist determines which points highlighted by CAD are clinically "significant".[88]

By 1995, a number of research groups in America and Europe were testing artificial neural networks (programs designed by different computer software companies) for analysing a variety of medical datasets. However, a survey of over 200 articles published between 1990 and 1995 which discussed ANN yielded from a Medline search reveals that none of them described an actual clinical trial using neural networks. Assessments of the uses of neural networks in computer-aided diagnosis remained centred on controlled trials of "textbook" cases. But even in these simulated settings, findings showed that the performance of neural networks varied considerably, alerting programmers and medical practitioners alike to the fact that the selection of data inputted to create an optimal network was no trivial task. "Despite claims made by software vendors", stated one author in the *Lancet*, "building and testing any decision aid takes great skill, and statistical support is essential".[89]

Rather than freeing up experts' time and facilitating complex decision making, these early findings suggested that more statistical calculations and diligence was required to make the systems work. One area that was singled out as being particularly problematic for machine learning was in the case of training a computer to see morphological features on a cervical cytology slide. As another article in the series in the *Lancet* which evaluated ANN reported, "Not only is there no absolute consensus on what features of a cell or of a cytological smear contribute to an assessment of normality but also it is unclear exactly where the bounds of such 'normality' begin and end. Histological images are so complex that their automated interpretation requires a highly adaptive mathematical procedure".[90]

One explanation for the variability in neural networks (and other computer-aided diagnosis systems) was the fact that no standards had been set for the data inputted to the computer which allowed it to "see" correctly and make appropriate recommendations to the practitioner. Ironically, this is analogous to the lack of standards for mammography film developing and viewing conditions that challenged radiologists' abilities until the Standards Act of 1992. Furthermore, when reducing the decision-making process to a rule-bound formula, it was unclear which, if any, features of a digitized image should be used to indicate the existence of an actual pathological problem. One author assessing neural networks pointed to two vital issues:

> Are there sufficient data to train the network and how should information about the problem domain be represented at the input of the network? … With images, raw pixel values by themselves have little or no information content; regions of interest can only be identified on the basis of features such as texture or contrast. The extraction of such features is essential to the success of a neural network application and knowledge about the problem domain is required in order to select features with high information content.[91]

While one problem with human observers was their variable skills, neural network performance was likewise degraded by the input of poor data. "Experience", it was stated, "is just as important for a neural network as it is for man".[92] This point was reiterated in another article that pointed out that "there is not one ANN, but an infinite number, and art, logic, and luck are all involved in selecting a useful one. The careful fitting, pruning, and general tweaking of an ANN when applied to a real problem requires just as much experience and training as any other analytical process. ANNs therefore, rather than removing the need for statistical modelling skills merely replace them with a different form of experience".[93]

These issues are central to the application of computer-aided diagnosis and neural networks to mammography since feature extraction, defining representational "normality", and assigning pixel values are essential to image interpretation. But in the process of standardizing software, datasets, and pathological features, new sets of skills and criteria of decision-making emerge which place new demands on the human experts for whom computer vision was intended to assist. While it is true that computers are not subject to boredom, fatigue, and distraction, it is becoming apparent that their proper design and function has instigated a professional transformation propelled by the introduction of new technologies which in the process of "stabilizing" continue to impact and redefine the training and skills of the medical profession for whom they were designed to assist.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous referees for their careful reading and suggestions for improvement.

REFERENCES

1. Stuart Blume, *Insight and industry: On the dynamics of technological change in medicine* (Cambridge, MA, 1992), 35.
2. Edward Tenner, *Why things bite back: Technology and the revenge of unintended consequences* (New York, 1996), 347.
3. Vassili Kovalev and Maria Petrou, "Texture analysis in three dimensions as a cue to medical diagnosis", in Isaac Bankman (ed.), *Handbook of medical imaging: Processing and analysis* (San Diego, 2000), 231.
4. D. B. Kopans, "Mammography and the normal breast", *Breast imaging* (Baltimore, 2007), chap. 12.
5. Caroline A. Jones and Peter Galison, *Picturing science, producing art* (New York, 1998); Martin Kemp, *Seen/unseen: Art, science and intuition from Leonardo to the Hubble telescope* (New York, 2006); Barbara Stafford, *Echo objects: The cognitive work of images* (Chicago, 2007).
6. Jutta Schickore, *The microscope and the eye: A history of reflections 1740–1870* (Chicago, 2007); Alberto Cambrosio and Peter Keating, "Of lymphocytes and pixels: The techno-visual production of cell populations", *Studies in the history and philosophy of the biological and biomedical sciences,* xxxi (2000), 233–70; Brian Dolan, "Pedagogy through print: James Sowerby, John Mawe and the problem of colour in early nineteenth century natural history illustration", *The British journal for the history of science,* xxxi (1998), 275–304. The pioneering and now classic volume by Michael Lynch and Steve Woolgar, *Representation in scientific practice* (Boston, 1990), is a natural starting point in studies of visualization in science.
7. Ian Hacking, *Representing and intervening: Introductory topics in the philosophy of natural science* (Cambridge, 1983); Nicholas Rasmussen, *Picture control: The electron microscope and the transformation of biology in America 1940–1960* (Palo Alto, 1999); Lorraine Daston and Peter Galison, *Objectivity* (New York, 2007).
8. Andrew Warwick, "X-rays as evidence in German orthopedic surgery 1895–1900", *Isis,* xcvi (2005), 1–24; Bernike Pasveer, "Depiction in medicine as a two-way affair: X-ray pictures and pulmonary tuberculosis in the early twentieth century", in Ilana Löwy (ed.), *Medical change: Historical and sociological studies of medical innovation* (Paris, 1993), 85–104.
9. Bjorn Hofmann, "The technological invention of disease", *Journal of medical ethics,* xxvii (2001), 10–19.
10. Robert Aronowitz, *Unnatural history: Breast cancer and American society* (Cambridge, 2007) provides a further historical context for attitudes toward breast cancer screening and treatment, but does not refer at all to debates about computer-aided diagnosis in mammography.
11. Barron Lerner, *The breast cancer wars: Fear, hope, and the pursuit of a cure in twentieth-century America* (Oxford, 2001); Aronowitz, *Unnatural history* (ref. 10).
12. American Cancer Society, "Proceedings of the workshop on cost of screening mammography", *Cancer*, lx (1987), 1669–702.
13. Barbara Monsees, "Screening mammography: Who will meet the need?", *Radiology*, clxxxiv (1992), 30–31.
14. Vicky Singleton and Mike Michael, "Actor-networks and ambivalence: General practitioners in the UK Cervical Screening Programme", *Social studies of science*, xxiii (1993), 227–64; Adele

Clarke and Monica Casper, "From simple technology to complex arena: Classification of pap smears, 1917–90", *Medical anthropology quarterly*, x (1996), 601–23.

15. Monsees, "Screening mammography" (ref. 13), 31.
16. Lerner, *Breast cancer wars* (ref. 11); Aronowitz, *Unnatural history* (ref. 10); H. Gilbert Welch, *Should I be tested for cancer* (Berkeley, 2004), for wider discussion of early-detection techniques in cancer screening.
17. A few works worth mentioning here provide relevant theoretical discussions that frame ways that 'calculation' is developed in science and technology studies to address apparent limitations of human expertise and establish better standards and protocols of practice: J. Rosser Matthews, *Quantification and the quest for medical certainty* (Princeton, 1995); Joel Howell, *Technology in the hospital: Transforming patient care in the early twentieth century* (Baltimore, 1995); Stanley Reiser, *Medicine and the reign of technology* (Cambridge, 1978).
18. Ferris Hall, "Screening mammography: Potential problems on the horizon", *New England journal of medicine*, cxxxiv (1986), 53–55.
19. Henry Collins, *Artificial experts: Social knowledge and intelligent systems* (Cambridge, MA, 1990); A. Kidd and M. Welbank, "Knowledge acquisition", in J Fox (ed.), *Expert systems 'State of the Art' report* (Oxford, 1984), 71–80.
20. Gerard Jorland, Annick Oppinel and George Weisz, *Body counts: Medical quantification in historical and sociological perspective* (Montreal, 2005); Peter Keating and Alberto Cambrosio, *Biomedical platforms: Realigning the normal and the pathological in late twentieth-century medicine* (Cambridge, MA, 2003); Harry Marks, *The progress of experiment: Science and therapeutic reform in the United States, 1900–1990* (Cambridge, 1997).
21. Marc Berg, *Rationalizing medical work: Decision support techniques and medical practices* (Cambridge, MA, 1997).
22. Berg, *Rationalizing medical work* (ref. 21), 27.
23. Berg, *Rationalizing medical work* (ref. 21), 30.
24. Berg, *Rationalizing medical work* (ref. 21), 37. Other conceptualizations of medicine as individual craft versus standardized science are in Warwick Anderson, "The reasoning of the strongest: The polemics of skill and science in medical diagnosis", *Social studies of science,* xxii (1992), 653–84.
25. Lorraine Daston and Peter Galison, "The image of objectivity", *Representations*, xl (1992), 81–128; Peter Galison, "Judgment against objectivity", in Caroline Jones and Peter Galison (eds), *Picturing science, producing art* (New York, 1998), 327–59.
26. Daston and Galison, "Image of objectivity" (ref. 25), 110.
27. Daston and Galison, "Image of objectivity" (ref. 25), 118.
28. Galison, "Judgment against objectivity" (ref. 25), 338.
29. Berg, *Rationalizing medical work* (ref. 21).
30. Elizabeth Krupinski, "The future of image perception in radiology: Synergy between humans and computers", *Academic radiology*, x (2003), 1–3, p. 1.
31. A. D. deGroot, "Perception and memory versus thought", in B. Kleinmuntz (ed.), *Problem solving research, methods and theory* (New York, 1966), 19–50.
32. W. G. Chase and H. A. Simon, "Perception in chess", *Cognitive psychology*, iv (1973), 55–81.
33. Anders Ericsson and Jacqui Smith, *Toward a general theory of expertise: Prospects and limits* (Cambridge, 1991).
34. Anders Ericsson, "How experts attain and maintain superior performance: Implications for the enhancement of skilled performance in older individuals", *Journal of aging and physical activity*, viii (2000), 346–52.
35. Ericsson, *op. cit.* (ref. 34).

36. C. F. Nodine and E. A. Krupinski, “Perceptual skill, radiology expertise and visual test performance with Nina and Waldo”, *Academic radiology*, v (1998), 603–12.
37. For an example of a current study investigating expert visualization, see E. Krupinski, A. Tillack, L. Richter, J. Henderson, A. Bhattacharyya, K. Scott, A. Graham, M. Descour, J. Davis, and R. Weinstein, “Eye movement study and human performance using telepathology virtual slides: Implications for education and differences with experience”, *Human pathology*, xxxvii (2006), 1543–56.
38. Alan Lesgold, Harriet Rubinson, Paul Feltovich, Robert Glaser, Dale Klopfer, and Yen Wang, “Expertise in a complex skill: Diagnosing X-ray pictures”, in Michelene Chi, Robert Glaser, and M. J. Farr (eds), *The nature of expertise* (Hillsdale, NJ, 1988), 311–42.
39. Lesgold *et al.*, “Expertise in a complex skill” (ref. 38), 329.
40. Lesgold *et al.*, “Expertise in a complex skill” (ref. 38), 319.
41. Lesgold *et al.*, “Expertise in a complex skill” (ref. 38), 337.
42. Lesgold *et al.*, “Expertise in a complex skill” (ref. 38), 340.
43. Lesgold *et al.*, “Expertise in a complex skill” (ref. 38), 319.
44. Lawrence Bassett, Viken Manjikian, and Richard Gold, “Mammography and breast cancer screening”, *Surgical clinics of North America*, lxx (1990), 775–800; S. A. Feig, G. D. Shaber, and A. Patchefsky, “Analysis of clinically occult and mammographically occult breast tumors”, *American journal of roentgenology*, cxxviii (1977), 403.
45. Lawrence Bassett, “Clinical image evaluation”, *Radiologic clinics of North America*, xxxiii (1995), 1027–39, p. 1032.
46. Philip Strax, “The role of the radiologist in mammography screening”, *Cancer*, lx (1987), 1675–7, p. 1675.
47. Americal College of Radiology, *Mammography control manual* (Reston, VA, 1990); Bassett, “Clinical image evaluation” (ref. 45), 1033.
48. Strax, “The role of the radiologist” (ref. 46), 1675.
49. Stephen Barley and Beth Bechky, “In the backrooms of science: The work of technicians in science labs”, *Work and occupations,* xxi (1994), 85–126; G. V. Larkin, “Medical dominance and control: Radiographers in the division of labor”, *Sociological review,* xxvi (1978), 843–58; Julian Orr, *Talking about machines: An ethnography of a modern job* (Ithaca, 1996); Lucy Suchman, “Making work visible”, *Communications of the ACM,* xxxviii (1995), 56–68; Lucy Suchman, *Human-machine reconfigurations* (Cambridge, 2007).
50. Florence Houn, Kathleen Franke, Maria Elliott, Charles Finder, Roger Burkhart, and Ruth Fischer, “The Mammography Quality Standards Act of 1992: History and process”, *Food and drug law journal*, l (1995), 485–92.
51. H. P. Chan, Kunio Doi, C. J. Vyborny, R. A. Schmidt, C. Metz, K. L. Lam, T. Ogura, Y. Wu, and G. MacMahon, “Improvement in radiologists’ detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis”, *Investigative radiology*, xxv (1995), 1105–10.
52. Quoted in Clarke and Casper, “From simple technology” (ref. 14), 607.
53. Chan *et al.*, “Improvement in radiologists’ detection” (ref. 51), 1109.
54. J. G. Elmore, C. K. Wells, C. H. Lee, D. H. Howard, and A. R. Feinstein, “Variability in radiologists’ interpretations of mammograms”, *New England journal of medicine*, cccxxxi (1994), 1493–9.
55. Elmore *et al.*, “Variability in radiologists’ interpretations” (ref. 54), 1493.
56. M. N. Linver, S. B. Paster, R. D. Rosenberg, C. R. Key, C. A. Stidley, and W. V. King, “Improvement in mammography interpretation skills in a community radiology practice after dedicated teaching courses: 2-year medical audit of 38,633 cases”, *Radiology*, clxxxiv (1992), 39–43.
57. Heang-Ping Chan, Kunio Doi, Simranjit Galhotra, Carl Vyborny, Heber MacMahon, and Peter Jokich,

"Image feature analysis and computer-aided diagnosis in digital radiography: 1: Automated detection of microcalcifications in mammography", *Medical physics*, xiv (1987), 538–48, p. 538.

58. Chan *et al.*, "Image feature analysis" (ref. 57), 539.
59. Chan *et al.*, "Image feature analysis" (ref. 57), 540.
60. Carl Vyborny and Maryellen Giger, "Computer vision and artificial intelligence in mammography", *American journal of roentgenology*, clxii (1994), 699–708, p. 701.
61. Casimir Kulikowski, "Artificial intelligence methods and systems for medical consultation", *IEEE transactions on pattern analysis and machine intelligence*, ii (1980), 464–76, p. 465.
62. Anthony Komaroff, "Algorithms and the 'art' of medicine", *American journal of public health*, lxxii (1982), 10–13, p. 10.
63. Edward Shortliffe, Bruce Buchanan, and Edward Feigenbaum, "Knowledge engineering for medical decision making: A review of computer-aided clinical decision aids", *Proceedings of the IEEE*, lxvii (1979), 1207–24.
64. N. Asada, K. Doi, and H. MacMahon, "Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung disease: Pilot study", *Radiology*, clxxvii (1990), 857–60, p. 857.
65. Asada *et al.*, "Usefulness of an artificial neural network" (ref. 64), 860.
66. Y. Wu, K. Doi, M. L. Giger, and R. M. Hishikawa, "Computerized detection of clustered microcalcifications in digital mammograms: Applications of artificial neural networks", *Medical physics*, xix (1992), 555–60; Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer", *Radiology*, clxxxvii (1993), 81–87.
67. Wu *et al.*, "Artificial neural networks in mammography" (ref. 66), 86.
68. P. Taylor, S. Hajnal, M. H. Dilhuydy, and B. Barreau, "Measuring image texture to separate 'difficult' from 'easy' mammograms", *The British journal of radiology*, lxvii (1994), 456–63, emphasis added.
69. H. L. Dreyfus and S. E. Dreyfus, "Peripheral vision", *Organization studies*, xxvi (2005), 779–92.
70. Edward Feigenbaum and Pamela McCorduck, *The fifth generation: Artificial intelligence and Japan's computer challenge to the world* (Lebanon, IN, 1983), 64.
71. Dreyfus and Dreyfus, "Peripheral vision" (ref. 69), 788.
72. Krupinski, "The future of image perception" (ref. 30), 2.
73. Krupinski, "The future of image perception" (ref. 30), 2.
74. Elizabeth Krupinski, "Response", *Radiology*, xi (2004), 616.
75. Taylor *et al.*, "Measuring image texture" (ref. 68).
76. For similar research on "easy" and "difficult" classification in Britain, see also Simon Cross, Joseph Downs, Pierre Dreszet, Z. Ma, and Robert Harrison, "Intelligent decision support systems in the cytodiagnosis of breast carcinoma", in Paulo Lisboa, Emmanuel Ifeachor, and Piotr Szczepaniak (eds), *Artificial neural networks in biomedicine* (London, 2000).
77. Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer aided diagnosis", *Academic radiology*, vi (1999), 22–23.
78. D. Signorini and Jim Slattery, "Neural networks", *Lancet*, cccxlvi (1995), 1500.
79. H. L. Dreyfus and S. E. Dreyfus, *Mind over machine: The power of human intuition and expertise in the era of the computer* (New York, 1986).
80. Strax, "The role of the radiologist" (ref. 46), 1675.
81. S. R. Dodds, "Neural networks", *Lancet*, cccxlvi (1995), 1500.
82. Jeremy Wyatt, "Nervous about artificial neural networks?", *Lancet*, ccxlvi (1995), 1175–6, p.

1176.

83. Vyborny and Giger, "Computer vision" (ref. 60), 705.
84. R. M. Nishikawa, R. A. Schmidt, J. Papaioannou, R. Osnis, R. A. Haldemann, M. L. Heusler, D. E. Giger, C. Wolverton, C. Comstock, and K. Doi, "Performance of a prototype clinical 'intelligent' mammography workstation", in K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt (eds), *Digital mammography '96* (Amsterdam, 1996), 93–96; M. D. Mugglestone, R. Lomaz, A. G. Gale, and A. R. M. Wilson, "The effect of prompting mammographic abnormalities on the human observer", *ibid*., 87–93.
85. Stuart Blume, *Insight and industry: On the dynamics of technological change in medicine* (Cambridge, MA, 1992), 35.
86. Michael Ulissey and Jimmy Roehrig, "Mammography — Computer-aided detection", downloaded from http://emedicine.medscape.com/article/346931-overview on 1 June 2009.
87. Berg, *Rationalization medical work* (ref. 21), 79.
88. E. Alberdi, A. A. Povyakalo, L. Strigini, P. Ayton, M. Hartswood, R. Proctor, and R. Slack, "Use of computer-aided detection (CAD) tools in screening mammography: A multidisciplinary investigation", *The British journal of radiology,* lxxviii (2005), S31–S40.
89. Wyatt, "Nervous" (ref. 82), 1177.
90. Richard Dybowski and Vanya Gant, "Artificial neural networks in pathology and medical laboratories", *Lancet*, cccxlvi (1995), 1203–7, p. 1203.
91. Lionel Tarassenko, "English as she is wrote", *Lancet*, cccxlvi (1995), 1712.
92. Dybowski and Gant, "Artificial neural networks" (ref. 90), 1204.
93. Signorini and Slattery, "Neural networks" (ref. 78), 1500.