A Review of Perceptual Expertise in Radiology-How it develops, How we can test it, and Why humans still matter in the era of Artificial Intelligence

Stephen Waite, MD, Zerwa Farooq, MD, Arkadij Grigorian, MD, Christopher Sistrom, MD, Srinivas Kolla, MD, Anthony Mancuso, MD, Susana Martinez-Conde, PhD, Robert G. Alexander, PhD, Alan Kantor, MD, Stephen L. Macknik, MD

As the first step in image interpretation is detection, an error in perception can prematurely end the diagnostic process leading to missed diagnoses. Because perceptual errors of this sort—"failure to detect"—are the most common interpretive error (and cause of litigation) in radiology, understanding the nature of perceptual expertise is essential in decreasing radiology's long-standing error rates. In this article, we review what constitutes a perceptual error, the existing models of radiologic image perception, the development of perceptual expertise and how it can be tested, perceptual learning methods in training radiologists, and why understanding perceptual expertise is still relevant in the era of artificial intelligence. Adding targeted interventions, such as perceptual learning, to existing teaching practices, has the potential to enhance expertise and reduce medical error.

Key Words: Visual perception; Expertise; Radiology; Visual search; Perceptual learning; Attention; Holistic processing; Gist; Artificial intelligence.

INTRODUCTION

Optimizing perceptual expertise in radiology is of critical importance for patient care. During radiologic interpretation, detection—noting that a potentially significant finding is present that merits further analysis—has primary importance, because all following steps leading to diagnosis rely on its efficacy (1). Despite continual efforts to optimize perception during radiologic interpretation, however, the error rate in radiological readings has not improved in the last 7 decades (2–4). This problem—compounded by increasing imaging volumes and examination complexity—mandates a deeper understanding of the nature of expertise to improve the training and accuracy of practicing clinicians.

In this review, we discuss the role of perceptual expertise in radiology, search strategies, and current training methods. We also discuss gist processing theories, ideal methods of perception testing, and why human perception is still relevant to radiology in an era of emerging artificial intelligence (AI).

Acad Radiol 2020; 27:26–38

From the Department of Radiology, SUNY Downstate Medical Center, 450 Clarkson Avenue, Brooklyn, NY 11203, USA (S.W., Z.F., A.G., S.K.); Departments of Ophthalmology, Neurology, and Physiology/Pharmacology, SUNY Downstate Medical Center, Brooklyn, New York (S.M.-C., R.G.A., S.L.M.); Department of Radiology, University of Florida College of Medicine, Gainesville, Florida (C.S., A.M.); Schneider Institutes for Health Policy, Brandeis University, Waltham, Massachusetts (C.S.); Department of Radiology, Lincoln Hospital—NYC Health and Hospitals, Bronx, New York (A.K.). Received July 18, 2019; revised August 26, 2019; accepted August 27, 2019. No funding sources were used for this manuscript. **Address correspondence to:** S.W. e-mail: waite.stephen@gmail.com

https://doi.org/10.1016/j.acra.2019.08.018

ERROR RATES IN RADIOLOGY

In 1949, Garland found that radiologists incurred an error rate of 33% in the interpretation of *positive* films (films that contain an abnormality), measured against the consensus of a group of experts (2). In a typical clinical practice (comprised of normal *and* abnormal studies), the diagnostic error rate has been found to approximate 4% (5), a rate that translates into approximately 40 million interpretive errors per year worldwide (6). Since Garland's pioneering studies, significant error rates have been noted in virtually all imaging modalities, including mammography, chest X-rays (CXR), skeletal X-rays, ultrasound, and CT and MRI of various organs, involving radiologists not only in private practice (5), but also in academic settings, where interpretive error rates ranges widely from 13% to as high as 90% depending on experimental conditions, imaging modality, and the definition of error (2,3,7−10).

What is a "Perceptual" Error?

Because of the subjective nature of radiologic interpretation, the definition of an "error" (vs. observer variation) is established by expert opinion (4). Thus, a conclusive error entails a substantial discrepancy with respect to peer consensus (4).

There are a couple of limitations to this definition. The inherently subjective nature of the definition of "error," and its' severity if deemed present, has led to challenges to peer review of radiological discrepancies (11,12). In addition, not all errors have clinical impact. Dependent on the scenario and modality, up to 32% of errors have been found to be of "major" significance, usually defined as an error that can change clinical management (13,14).

Although radiologic errors can be classified in a number of ways (15), two broad categories of interpretive error are usually identified: cognitive and perceptual (6). Cognitive errors occur when a correct positive finding is followed by misclassification due to faulty reasoning or lack of knowledge (16). Omission, or false negative errors, occur when a radiologist fails to detect a perceptible lesion. In practice (and in this review) all omission errors are termed "perceptual" (6) (although it is important to note that this differs from how perceptual errors were originally defined by Kundel et al. using fixation times (17); see Section 10).

Perceptual errors are the most important type of error in radiology, accounting for 42%−80% of all interpretive errors (15,16,18,19). They are also the most common reason for malpractice suits against radiologists, comprising 78% of all cases according to one study (20).

EYE MOVEMENTS AND SEARCH PATTERNS IN RADIOLOGY

Eye Movements and Expertise

When scanning the environment, the eyes make jerky movements called "saccades," interleaved with fixation periods (21). Saccades serve to point the fovea (the central part of the retina, which has sufficient photoreceptor density to provide high-resolution vision) to areas of interest in an image, and thus capture detailed snapshots of such locations (Fig 1).

Expert radiologists generally fixate on abnormalities faster than novices. Further, total image search time decreases with increasing levels of expertise (22). Experts also produce fewer total fixations than novices. These differences may be due to novices spending more time looking at irrelevant but salient structures (such as the heart on a CXR, when the lungs are more important to analyze in a nodule detection task), and to experts demonstrating more effective search strategies (22,23) (Fig 2).

Although most studies have examined the effects of expertise on plain film (2D) interpretation, expert radiologists are also more accurate and faster than novices during interpretation of volumetric imaging, such as CT and MRI (24−26).

Search Patterns During Plain Film Interpretation

As a practical matter, research findings concerning human perception of medical images have not been translated into heuristics that improve training. Although there are published guidelines on how to interpret various radiologic examinations (i.e., chest X-ray interpretation as in (27)), few studies have demonstrated their efficacy and when vigorously analyzed, most educational tools have had mixed results.

For example, novices are thought to benefit from an orderly and comprehensive search pattern, so-called "systematic viewing" (28) (Fig 2). Conceptually, systematic viewing would help readers achieve more complete coverage of the image, and thereby reduce the number of overlooked abnormalities (29).

Evidence to support the value of systematic viewing is wanting, however. Whereas Van Geel et al. found that students trained in systematic viewing methods inspected a larger portion of images than untrained students (30), both groups performed comparably in chest radiographic interpretation. Kok et al. arrived to similar findings (29). Thus, the available data indicate that an emphasis on systematic viewing may not be justified.

One potential reason that predefined search patterns fail to consistently improve accuracy may be that experts themselves do not read plain films in a consistent, standardized manner. Instead, experts tend to use a variety of nonsystematic search patterns, so-called "free search," when looking at plain film images. Their eye movements appear more affected by the findings on the radiograph than by any preplanned search pattern (Fig 2) (31). In practice, consistent search patterns might be *detrimental* to accurate diagnosis (29,32).

Search Patterns in Volumetric Imaging

Relatively little is known about search strategies employed during interpretation of volumetric imaging (24). During stack mode viewing, radiologists simulate motion by scrolling through sequential images, searching for suddenly appearing lesions that stand out from the background (24). Thus, the fundamental characteristics found in the search of static images do not necessarily apply to volumetric search.

A few studies have begun to shed light on how radiologists search volumetric examinations. Venjakob and den Boaer et al. (33,34) defined different types of scroll behavior during CT interpretation and temporally related scroll movements to cognitive data via a think-aloud strategy, whereupon radiology residents verbalized their thoughts while reading CT scans (34). For example, half runs and

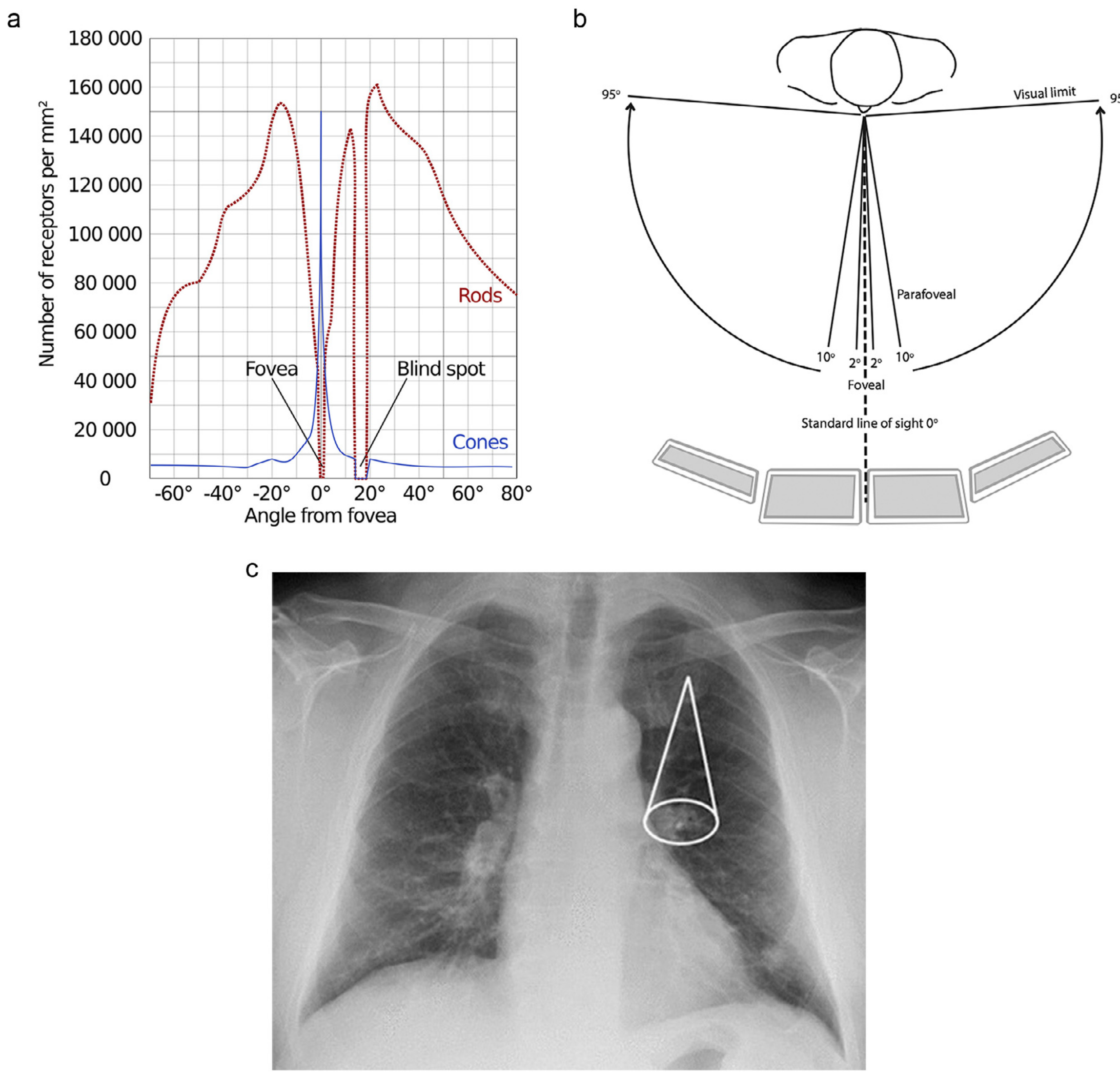

Figure 1. (a) and (b). Reprinted from (111) with permission. (a) Relationship of rod and cone density to the distance from the fovea. The retina contains two different types of photoreceptors, rods and cones. The region of the retina with the highest visual acuity is the area of highest cone density, the fovea centralis. (b) Horizontal field of view of the human eye. The fovea is the portion of the retina with the highest spatial resolution, constituting the central 2°–4° of the visual field. (c) Reprinted from (112) with permission. Radiograph illustrating the useful visual field on a chest radiograph (CXR) that can be processed with high-resolution foveal vision as the observer moves his or her eyes around an image to gather information. (Color version of figure is available online.)

oscillations ("local" movements covering less than 50% of the stack slices) were often associated with analysis (i.e., cognitive activities including the characterization of findings) (35).

Drew et al. identified two different global strategies adopted by radiologists during a nodule detection task in chest CT. "Scanners" searched each slice widely, before moving on to the next depth. "Drillers" held their eyes relatively still in the *x* and *y* plane, limiting their search to a single lung quadrant while quickly scrolling—drilling—through slices in the *z*-axis (volumetric depth) (36). Kelahan et al. found this categorization imperfect when applied to CT scans of the abdomen and pelvis, however (37).

Differences Between 2D and Volumetric Search

Radiologists are generally taught to read volumetric imaging in a systematic, often organ based, manner. This teaching may be more efficacious compared with analogous instructions for plain film imaging, as readers must view hundreds or even thousands of images (38) during cross-sectional interpretation. Given the volume of images, a structured approach may be helpful in focusing attention to specific regions/organs (i.e., the liver), thereby avoiding perceptual errors.

Because of the large amount of data inherent to volumetric imaging, radiologists cannot exhaustively foveate all regions of interest, but must rely, at least partly, on detecting signals

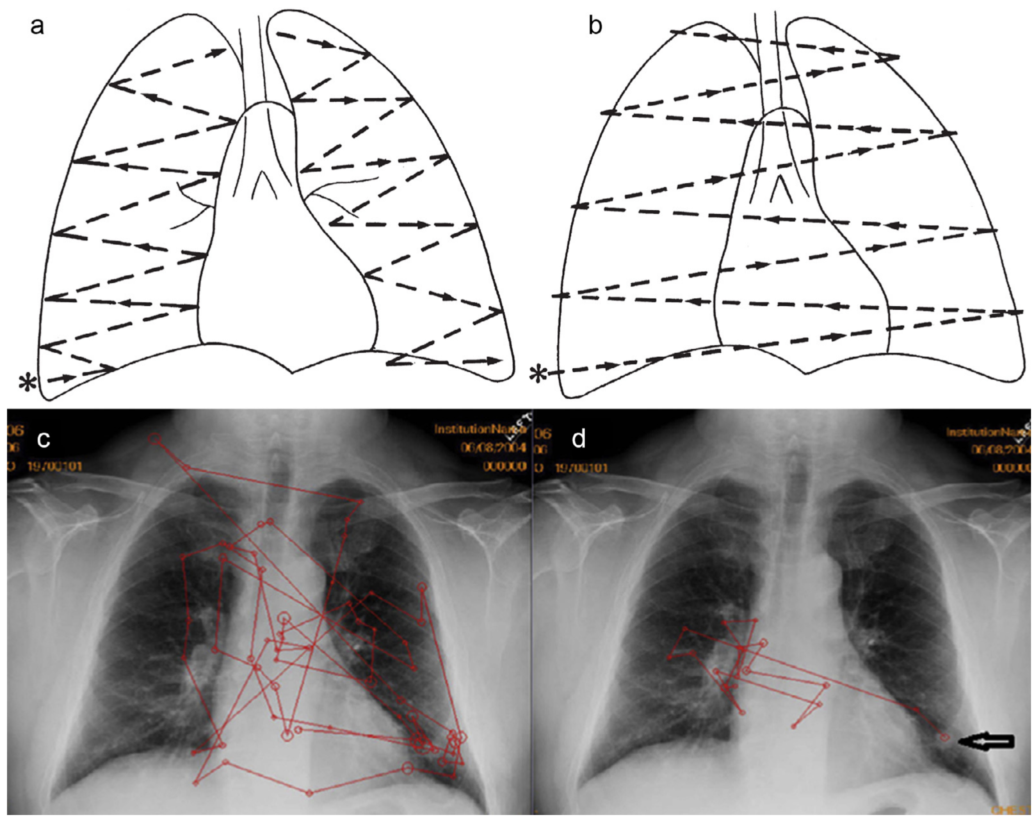

Figure 2. Reprinted from (28) with permission. (a) This commonly taught search pattern for examination of the lungs during CXR interpretation involves starting at the right base (*) (the costophrenic angle) and examining the right lung and then the left lung. (b) A second look is then performed to compare the right and left lungs, as bilateral symmetry is assumed to be useful in recognizing abnormalities (113). (c and d) Reprinted from (111) with permission. Typical scanpaths of a novice (c) and an expert (d) radiologist, both searching a CXR which has a nodule at the left base (arrow in [d]). This free search pattern (d) is typically employed by experts and differs from the formal radiologic training given in residency. Instead, it indicates the flexible use of search strategies as a function of immediate visual information. The expert radiologist (d) has more efficient scanpaths (red lines) than the novice (c), with fewer fixations (circles), less coverage of the image, fewer saccades, and faster arrival at the abnormality. (Color version of figure is available online.)

with their peripheral vision (39). As an example, Miller et al. surmise that secondary to the presence of hundreds of subsegmental pulmonary arteries, radiologists cannot directly inspect them all when excluding small pulmonary emboli. Rather, they rely on their peripheral vision with resultant low accuracy (40). Similarly, during CT nodule detection tasks, radiologists foveate only between roughly 27% and 69% of lung tissue; a large amount of the parenchyma is never examined with foveal vision (36,41).

There are notable limitations to peripheral vision. As peripheral vision cannot provide the kind of fine spatial discriminations that characterizes foveal vision, detectability of certain lesions differ in 3D vs. 2D image searches (39). Eckstein et al. found higher detectability for calcifications in single slice images and relatively improved detection of masses in volumetric imaging (39). Using saliency maps, Wen et al. demonstrated that the great majority of radiologists (42) use dynamic (motion based) information when interpreting cross-sectional imaging (43). Both Wen and Eckstein studies lend support to the notion that observer performance in 2D search might not generalize to volumetric search tasks (39,43), as the 2 types of search entail different perceptual tasks.

Optimal Search Strategies

The optimal viewing strategy for any given imaging modality remains unknown. Referring to this gap in knowledge, Van der Gijp et al. (22) noted that studies conducted over the last 2 decades focused mainly on differences between experts and novices, while comparatively neglecting theory-driven initiatives to *improve* detection. Thus, there is a need for the field to develop more efficient strategies and methods to accelerate and advance training. Importantly, any "optimal" interpretive techniques will be different for 2D vs. 3D examinations, are examination specific, and may change during interpretation.

With regard to 2D imaging, systematic viewing has been generally advocated for CXR interpretation, though not shown to improve diagnostic performance (32,44). Similarly, taught search strategies for other 2D examinations, including skeletal X-rays, abdominal X-rays, and mammography, are unprincipled. Moreover, because radiology experts have limited ability to accurately report on their viewing behavior, what they advance as their search patterns may not reflect their actual eye movements (i.e., as measured with eye tracking methodology) (45).

Optimal strategies for *3D* imaging interpretation such as CT is likely modality and region specific. For example, rather than demonstrating a clear preference for either drilling or scanning (as during the nodule detection task (36)), radiologists both vigorously drill *and* extensively scan during interpretation of digital breast tomosynthesis (46). In addition, strategies may change *during* interpretation. Thus, even if a drilling strategy is ideal for pulmonary nodule detection, it may not be efficient for examination of the mediastinum on the same CT scan.

The abovementioned challenge to creating optimal search strategies suggest that future improvements to perceptual expertise training may lie in determining the precise nature of radiologic abnormality perception. Once we understand precisely what expert radiologists are looking for, we may be able to optimize training regimes.

PREEXISTING VISUOSPATIAL SKILLS AND RADIOLOGIC EXPERTISE

Measuring the perceptual abilities of radiology applicants and residents would be of great practical importance, as trainees are usually selected on the basis of their academic records, letters of recommendation, and interviews, none of which directly pertain to perceptual abilities. The existing model of training assumes that almost all trainees will eventually reach an acceptable standard with practice and semantic knowledge. However, it is possible that trainees with higher preexisting skills (i.e., visual-spatial processing) may reach a *higher level* of expertise—or may achieve the highest level of expertise *more quickly*—than trainees with lower preexisting skills (47). A relevant perceptual test might therefore help determine how much individual residents may benefit from training, and, ultimately, how they will perform as radiologists (48).

To this end, investigators have attempted to identify perceptual requirements for both learning and practicing radiology, as well as ascertain whether practicing radiologists have superior perceptual skills outside of imaging. Extant evidence suggests that radiologic expertise is a learned, task-specific, skill, and that expertise in visual search and/or perceptual discrimination does not carry over to nonradiological tasks (49,50). Although some preliminary experiments demonstrated initial promise (51), no visuospatial ability test currently exists to determine whether someone is likely to become an "expert" radiologist. Ongoing research aims to determine the role of any relevant perceptual abilities which may exist pretraining (48).

HOLISTIC "GIST" PROCESSING THEORY

Prevailing models of medical image perception are based on the premise that expert observers process a medical image holistically at the first glimpse. The oldest of such models, the global-focal model (52,53), posits that medical experts rapidly extract a global impression of an image. This impression consists of a comparison between the contents of the image and the expert's prior knowledge about the appearance of normal and abnormal medical images (i.e., the expert's schemata). This process enables experts to identify perturbations (deviations from their schemata that indicate possible abnormalities) and direct their eyes toward their corresponding locations for further (i.e., foveal) examination (53,54). Features are subsequently scrutinized and tested against schemata to determine whether a finding is suspicious, in which case diagnostic decisions are made (55). Radiologists then either direct their gaze to additional suspicious locations based on information from the global impression, or engage in "discovery scanning" of the image (i.e., coarse screening of the image, conducted in order to detect other potential targets (56)). The global-focal model suggests that this procedure may be recursive: if a decision is not made after foveal examination, a new global impression can form, followed by a new discovery search. The more recent iteration of this model—the "holistic model"—posits that fast holistic processing can work in *parallel* with slower discovery scanning (54,56,57).

One of the global-focal and holistic models' principal predictions is that rapid initial global processing constrains subsequent search to suspicious areas in an image (58). This strategy may be available to experts but not novices, explaining why expert observers search medical images with higher efficiency—finding more abnormalities in a shorter timeframe, and with fewer eye movements—than novices do. Support for this hypothesis comes from studies showing that expert radiologists can identify subtle abnormalities on mammography and chest radiography displayed for only 250 milliseconds (54,55,58−61).

Another popular model posits that initial global processing (consisting of "global image statistics" like average orientation and average size of objects) signals if there is an abnormality (establishing its likelihood) *without* providing location information or constraining the subsequent serial search. Searchers can then change their strategy to a slower, more complete search for abnormalities (62−65) (Fig 3).

In addition, search may be faster for expert radiologists because of "visual chunking" of information across the image. If so, task-relevant information could be processed by experts as "chunks" or "units," instead of as individual pieces of information, reducing costs to attention and working memory (66−68).

Although the holistic processing theory is popular, one should note that the studies supporting it were conducted with plain film imaging, such as CXR (59,61) and mammography (60,65). However, the nature of volumetric imaging is such that no single image can provide meaningful global image statistics, or afford knowledge of image perturbations, throughout the entire dataset. Therefore, there is no "global signal" that can be extracted at any single point in time to either organize subsequent fixations or contribute to the reader's conviction that a subsequent search will uncover an abnormality (69). Thus, the holistic processing theory is incomplete in regard to volumetric imaging. In addition, recent Flash Preview Moving Window (FPMW) experiments, drawing from both "flash" methodology and "moving window" paradigms, fail to support the idea that processing the initial glimpse of a scene is beneficial to performance (70).

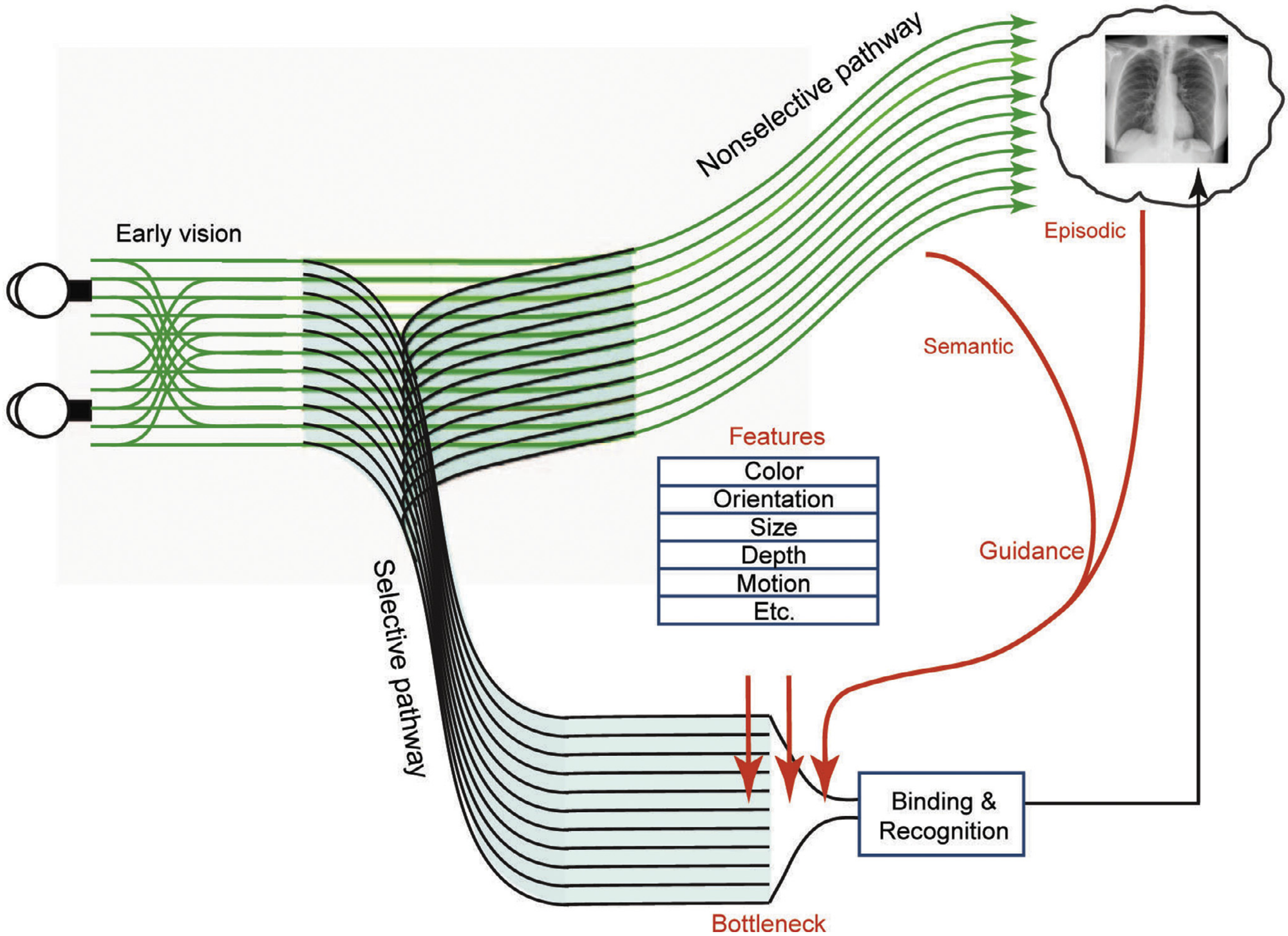

Figure 3. Figure modified from (114) with permission. Two-pathway architecture for visual processing. The selective pathway can bind features and recognize objects, but it is capacity limited. At its bottleneck, preference for further processing is given to items with certain basic attributes (such as color, orientation, and size), when those attributes match the appearance of a target object. However, these attributes do not fully explain the efficiency of search in the real world, where elements are arranged in a rule-governed manner—for example, people generally appear on horizontal surfaces. The regularity of scenes provides two kinds of scene-based guidance—semantic guidance, referring to the knowledge of the probability of the presence of an object in a scene and its probable location, and episodic guidance, referring to the memory of a *specific* previously encountered scene. In conjunction with the selective pathway, the nonselective pathway extracts statistics (such as velocity, direction of motion, and size) rapidly from the entire image. Although the nonselective pathway does not support precise object recognition, it provides information used in scene-based guidance to direct attention to important locations (such as the probable locations of nodules on CXR's). The visual experience is comprised of the products of both pathways (114). (Color version of figure is available online.)

THE DEVELOPMENT OF PERCEPTUAL EXPERTISE IN RADIOLOGY

Kelly et al. found that certain ocular metrics (such as time to first fixation) improved at an earlier stage of training than diagnostic accuracy in pneumothorax detection, and then plateaued before the end of formal training (71). In addition, Kok et al. found that attendings and medical students demonstrated similar viewing patterns, as determined by a ratio of long to short saccades, despite significant differences in accuracy during interpretation of diffuse vs. focal diseases on CXR (72).

Ravesloot et al. found that scores from image interpretation questions improved faster than knowledge-based questions (text-based factual questions) for the first 3 years of residency when residents took the Dutch Radiology Progress Test, a mandatory semiannual test taken by all Dutch radiology residents (73) (Figure 4). Using the same dataset, Rutgers et al. found that the 5-year development of resident image interpretation scores were comparable for 2D *and 3D* imaging, though junior residents had slightly lower interpretation scores on volumetric imaging (a small but significant finding) (74).

Thus, the combined evidence from both eye tracking and performance metrics (on image interpretation-based questions) indicate that the perceptual aspects of image interpretation develop before the ability to correctly interpret abnormalities or integrate them into a correct diagnosis. That is, perceptual skills begin to grow from the start of exposure to imaging, and radiology-specific factual knowledge contributes little to this initial development (73).

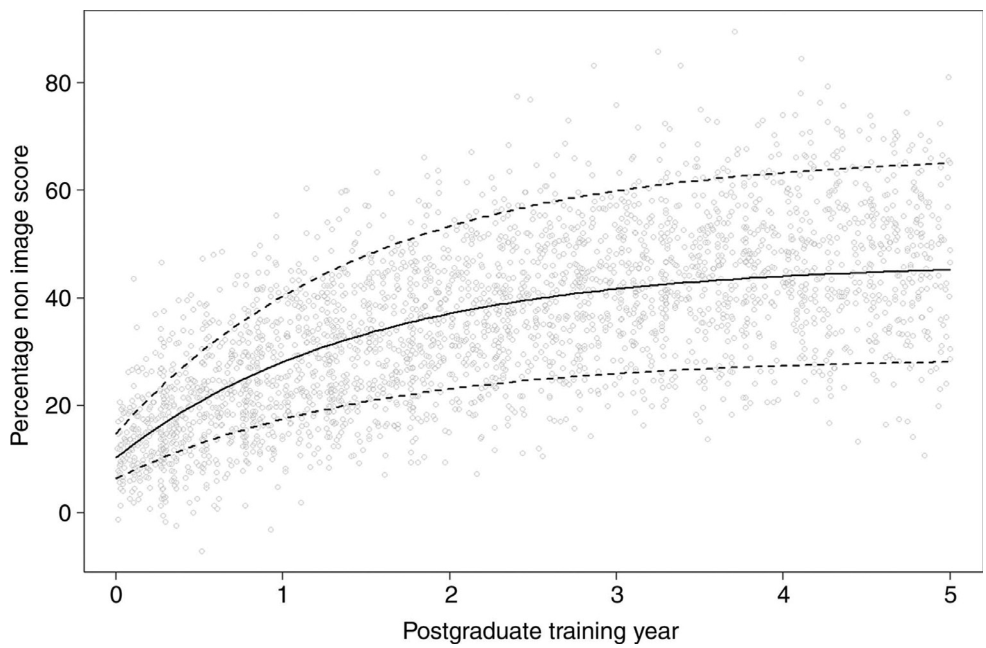

Figure 4. Reprinted from (73) with permission. Graph estimating image interpretation skill development during residency, as measured by the Dutch Radiology Progress test. Image score measures performance on image interpretation skills and represents percentage of the maximum possible score. It is calculated by subtracting the number of incorrect answers from the number of correct answers to account for guessing (making a negative value possible). The slope represents the speed of skill development and measures 16.8% during the first year of training. The slope decreases by 50% every year until it reaches 2.0% at the end of training. Note that the maximum image-score is estimated at 55.8%. Dotted lines represent the middle 95% of performances (73).

The above studies also suggest that perceptual expertise plateaus with a high level of error. Importantly, this plateau likely persists beyond the formal training period, an important issue for residency programs to address.

A WAY FORWARD—PERCEPTUAL LEARNING

One of the major problems in radiology education is the lack of formalization and verbalization of what exactly happens during visual information extraction (i.e., "How do you teach to see a nodule?") (75). Instead, it is assumed that advanced pattern recognition and automaticity eventually arise from long apprenticeships, leaving crucial aspects of learning to occur in an unsystematic fashion, over an unspecified time, with unquantified results. In 1983, Kundel and Nodine suggested that visual training, using typical examples of abnormalities and normal variants, would facilitate focused perceptual learning, whereupon trainees would learn to visually recognize abnormalities rather than to interpret medical imaging findings based on a formal set of explicit rules (76,77). However, little research to date has tackled the application of perceptual learning to radiology (77).

Sowden et al. conducted one of the first perceptual learning studies in radiologic imaging. They found that novice film readers improved their discriminations of clusters of microcalcifications in mammograms, and reduced their decision speeds, after a perceptual learning regime where they viewed 60 images, 3 times each day, for 4 days. Negative feedback was provided in the form of a computer beep when the wrong cluster location was selected (78). Remarkably, this work showed that, whereas radiologists in training may have already seen thousands of images, even small amounts of practice in a relatively short interval can produce significant improvements in sensitivity (78). More recently, Chen et al. examined the efficacy of perceptual learning on the performance of novices (with no prior knowledge of plain film interpretation) on the detection of hip fractures. They found that top performing novices achieved comparable accuracy to that of board-certified radiologists after training on 1280 images for 52 minutes (77). In a follow-up study using the same dataset, Adams et al. demonstrated that the detection accuracy of the top performing medically-naïve individuals for detecting femoral neck fractures on radiographs was comparable to that of the deep convolutional neural network GoogleNet (90.5% for humans vs. 94.4% with GoogleNet) after training on 640 images for less than 1 hour (79).

Whereas it is not known to what degree these findings might extrapolate to different pathologies or imaging modalities, the data thus far speak to the untapped potential of perceptual learning in radiology training (77) (Fig 5).

TESTING PERCEPTUAL EXPERTISE

Various efforts have been made over the years to accurately measure resident performance, including perceptual competence. Via the certification process, the American Board of Radiology attests to a certain level of achievement and function of its diplomates (80). Currently, the American Board of Radiology administers two examinations for residents to gain

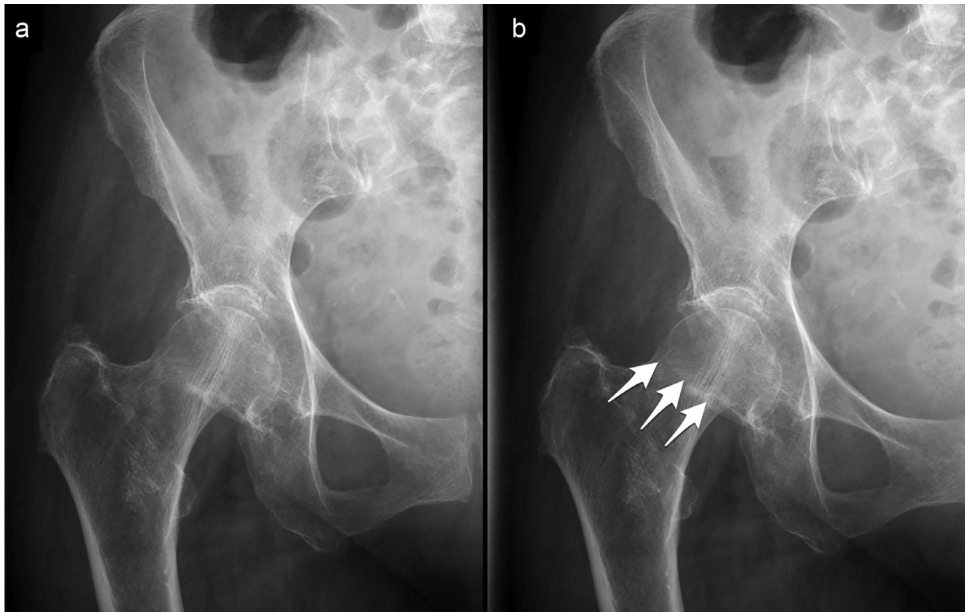

Figure 5. Reprinted from (77) with permission. (a) Example of an image shown during perceptual training of hip fracture identification. (b) Arrows represent the feedback provided in case of a wrong answer. Top novices achieved expert level accuracy in hip fracture detection in under 1 hour of perceptual training (77).

certification in diagnostic radiology—the core examination taken after 36 months of residency, and the certifying examination taken 15 months after the completion of training (81,82). These examinations are considered a final step in determining whether trainees meet current standards and are equipped for independent practice (80,83). Both examinations are computer-based and largely multiple choice (as of 2017 there are small numbers of fill-in-the blank and point-and-click questions (84)), replacing the previous written and in-person oral board examinations (81).

The validity of the core examination has been questioned and debated (82,85−87), with one major critique being that the multiple-choice format renders it an inauthentic representation of radiologic practice. Supporters readily acknowledge that the core examination does not mimic clinical practice, but argue that that is not the intent of the examination (88). Lack of ecological validity notwithstanding, the multiple-choice nature of both examinations fundamentally renders them inadequate for testing *perceptual* expertise.

Because normal images are generally not shown, false positive rates cannot be ascertained. Although less discussed in the literature, false positive errors are important to recognize, as they can lead to patient anxiety as well as unnecessary studies and procedures (89). Additionally, full datasets are not available for review. Lastly, in a multiple-choice format question, the candidate knows that one of the answer choices must be correct.

These factors limit the test's ability to ascertain whether the candidate would note important salient findings in a full dataset, or format a reasonable differential diagnosis when findings are made. Rather than testing the candidate's ability to organize observations into meaningful patterns and consider plausible diagnoses, multiple-choice items limit differential possibilities (Fig 6).

Sports scouts watch potential athletes on the field/court, understanding that the best way to select for expertise is to evaluate a player's ability to perform relevant task(s). Similarly, because detection is the first step in image interpretation (1), one way to test perceptual expertise would be to examine the ability of residents to interpret studies in a "realistic" scenario. Since 2011, faculty members of the University of Florida Radiology department use a critical care radiology evaluation that simulates an 8 hour in-house after hours (i.e., on-call) rotation, to test residents' interpretative abilities. Sixty-five cases (including normal studies) are presented on a "worklist" in random order over 8 hours: trainees are charged with creating brief, but cogent and complete, interpretations for all of them. Each case consists of an abstracted clinical scenario and deidentified images in full Digital Imaging and Communications in Medicine resolution garnered from examinations performed on real patients in nonambulatory setting(s) during routine clinical care (90). In their reports, residents are asked to identify the relevant abnormalities, pertinent negatives, and additional findings, as well as to indicate how urgently they would communicate their findings to the referring physician. Responses are scored by a trained cadre of attending radiologists using a grading key consisting of discrete examination findings, true/false assertions, and point weights. This method of evaluation has the advantage of simultaneously assessing perceptual expertise, critical thinking, problem solving, and consulting skills, similar to the previous in-person oral examinations. In addition, it avoids many pitfalls of the original testing rubric, including stress provoked from the oral milieu, potential subjectivity of examiners, presentation of selected scan sections instead of the full three-dimensional data set, inability to manipulate and postprocess images, and lack of normal studies (81). The data obtained from nationwide simulations would be an important quality control tool, providing heretofore unavailable data regarding resident progression throughout training, resident and *institutional* deficiencies, error types per modality, common perceptual deficiencies, and even the role of fatigue during examination sessions.

PERCEPTUAL EXPERTISE IN THE ERA OF MACHINE LEARNING

The exponential development and initial implementation of AI algorithms, particularly convolutional neural networks (CNN), toward radiological image interpretation, has led to a widespread concern within the field of radiology that diagnostic radiologists may lose relevance to the practice of medicine (91). Indeed, in 2016, Dr. Geoffrey Hinton, a prominent figure of deep learning and AI research, stated that within a timeframe of 5−10 years, radiology as a specialty will become obsolete due to exponential improvement of deep learning medical image analysis software (92). Consequently, one might conclude that training optimization and the accurate assessment of perceptual expertise will become moot issues in the near future.

We disagree with these arguments. Whereas many studies have shown CNNs to be successful in the detection of certain narrowly defined imaging features in two dimensional datasets (93,94), and more recently in three dimensional volume datasets (95), vast limitations and challenges remain, which make the replacement of human radiologists improbable for at least several decades. Lack of standardized training datasets, poor interoperability standards, inability to access relevant clinical information, regional variances in image acquisition/processing, and delayed adoption of technological advances in healthcare industry, are all serious obstacles for radiology AI (96). Another pertinent concern regarding widespread use of AI for image interpretation without supervision includes adversarial perturbation of these algorithms, which can have devastating consequences for patient care (97). Finally, the medicolegal aspects of AI in medicine are not well defined.

Although AI is finding its application in a variety of clinical dimensions and will undoubtedly be a useful adjunct, radiologists will still be necessary to commit to the interpretation of medical images for the foreseeable future. The concept of "*augmented* intelligence" has recently been introduced, whereupon radiologists capitalize on their ability not to detect or classify images/disease, but to make clinical judgements about the data (98). Rather than dismissing this technology, radiologists are encouraged to work with AI in a form of collective intelligence (98).

It should be noted that even in those aspects of human endeavor that have been outstripped by AI, there has been a resurgence of human relevance, indicating that hybrid systems in which humans collaborate with computers to solve problems can be superior in performance to either AI or humans alone (99). For example, in Freestyle chess, humans compete in tandem with computers, generally producing better results than either type of player in isolation, despite the fact that AIs alone have beaten humans alone at chess for decades (100). Similarly, radiologists aided by CNN-based algorithms may produce faster and more accurate diagnoses than either experts or deep learning algorithms alone—

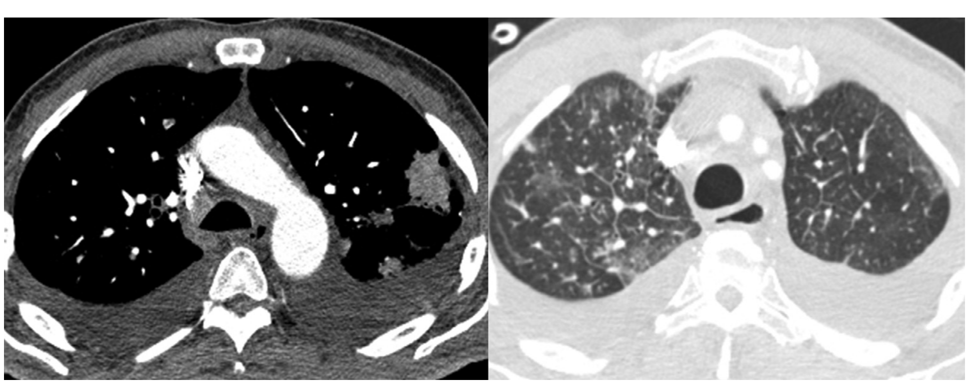

Figure 6. An example of a multiple-choice question as may be presented during the ABR core-examination. What is the most likely diagnosis given the images above?

a. Hypersensitivity pneumonitis
b. Fungal infection
c. Septic emboli
d. Lymphangitic carcinomatosis

Septal and bronchial wall thickening is identified in conjunction with a parenchymal mass in the left upper lobe. In addition, there are bilateral pleural effusions. Given the constellation of findings, the most likely answer is lymphangitic carcinomatosis. Providing the examinee with representative images precludes assessment of whether they would note salient findings if given the *entire* dataset. Further, given the list of possible answers, the tester knows that the study is abnormal *and* that there are only 4 possibilities. This construct both constrains the trainees' differential diagnosis (i.e., what if the trainee thought the findings represented community acquired pneumonia?) *and* limits assessment of the trainees' ability to call a study negative. Secondary to these shortcomings, perceptual expertise is not accurately assessed.

despite the incipient dominance of AIs in some radiological exams when compared to human performance. Thus, we predict a future where radiologists are mandatory as component human authorities, and their perceptual and decision-making skills are integrated with active supervision of AI tools (4,101).

Contrary to the notion that AI will displace radiologists, Duong et al. (102) propose "AI-empowered education," advancing how AI can compensate for deficiencies of the current apprenticeship model (103). AI applications can function as an "intelligent tutor," personalizing learning via tracking of resident competencies and reinforcement of challenging topics. For example, AI can preliminary interpret a case and assign it to a trainee whose profile indicates potential benefit. These algorithms can then direct trainees to review example case reports with similar radiologic features in conjunction with relevant literature. After discussion and attending review, this case can be added to a teaching file and the trainee's competency profile can be updated accordingly (102). This "live teaching file cataloging" can be extremely useful in generating a large database and improving the diversity of cases residents encounter. Increased case exposure and volume with associated feedback can promote expertise.

Zhang et al. (104) describe an adaptive computer aided education system which predicts the likelihood of a trainee missing a lesion on mammography based on the trainee's prior performance and imaging features of the lesion including tissue intensity, size, location, similarity to neighboring regions, and symmetry with the contralateral side. The average area under the Receiver operating characteristic curve for the described classifier was 0.607, demonstrating that it was able to predict which masses were detected and which were missed better than chance. With continued improvements to such algorithms and growth in the numbers and complexity of features that are amenable to extraction, AI could present cases to residents based on their specific perceptual "profile" and tendency to miss lesions with certain visual features. As expert radiologists do not have to *assign* features to abnormalities, such models afford for large databases of studies to be searched and educationally useful cases quickly identified (104). This methodology can also be useful in generating individualized textures for use in perceptual learning heuristics, further promoting perceptual expertise (69).

CAVEATS TO OUR UNDERSTANDING OF PERCEPTUAL EXPERTISE

When attempting to understand perception in radiologists, it is important to recognize that the ability of radiologists to detect abnormalities is influenced by numerous "extrinsic" factors, including clinical history, lesion prevalence and prevalence *expectations* (reviewed in (4)), environmental factors (such as distractions and viewing conditions), as well as observer knowledge and variance amongst radiologists.

Although perceptual errors are "...deemed to have occurred when an abnormality is retrospectively determined to have been present on a diagnostic image but was not seen by the interpreting radiologist at the time of primary interpretation" (6) we note that an important set of omission errors are caused by lack of reader knowledge, and therefore *cognitive* (6) in etiology. Radiologists approach film reading with information derived from learning and experience. Object knowledge represents understanding of the visual aspects of the object of search. In order for an abnormality to be recognized, the observer must be aware in

Figure 7. Lateral femoral notch sign in a 17-year old. Impacted lateral condylopatellar sulcus (arrow) is seen with an effusion (*). This finding is associated with an underlying Anterior Cruciate Ligament tear and frequently missed on radiographs. This example was missed by several radiologists during film review despite unlimited viewing times and knowing that the film was abnormal. Although eye tracking was not performed, finding is likely secondary to decision error, as knowledge of this sign is essential in noting the abnormality (115).

advance of the features (e.g., size, shape, and density) that make the object distinctive (105). Lending credence to the radiologic saying "You see what you know," if the reader does not have the background knowledge that certain imaging findings are abnormal, they will often not recognize those features as such. Decision (or "decision-making") errors, which are considered cognitive in nature (17), occur when a radiologist fixates on a lesion for a long period of time (over 0.5−1 second), but either does not consciously recognize the features or actively dismisses them (106,107). These errors likely include many cases where radiologists spend an unbounded time looking at an examination but fail to note salient abnormalities (Fig 7). Approximately 40% of omission errors in CXR (105), musculoskeletal plain films (108), and mammograms (109) have been established to be secondary to decision errors, indicating that the *contemporary* definition of perceptual error (usually determined without eye tracking) is often incorrect.

There is interobserver variability in radiologists' visual attention deployment. Wen et al. evaluated 16 saliency models in terms of how well they agreed with radiologists' eye positions during interpretation of CXR, CT scans, and positron emission tomography scans. They found that certain saliency models performed better for some radiologists, implying that different image information (i.e., intensity, orientation, edges, etc.) may be utilized in visual searches conducted by different individuals (110). How radiologists interpret examinations is therefore at least partially secondary to training and individual idiosyncrasies. This understanding could potentially help identify efficient strategies of attention deployment to improve diagnostic accuracy and training (110).

These factors must be considered when developing methods to decrease perceptual error, as they suggest a need for training individualization. AI may be a helpful adjunct in standardizing such elements.

CONCLUSION

There is no lack of instructional materials on how to make differential diagnoses for problematic findings, but such resources fail to address the first step of interpretation: perception. The fact that observational errors constitute the bulk of interpretive error in radiology, and that error rates have not changed in over half a century (4), highlights the need for new educational methods and the reassessment of present didactic and question-and-answer instruction techniques.

In this review we discussed the extant literature and propose that radiology's error rate has been recalcitrant to improvement secondary to an incomplete understanding of mechanisms underlying perceptual expertise. The ability of a radiologist to see abnormalities largely depends on their skill to recognize subtle shapes and textures embedded in a noisy background. Radiologic expertise may constitute the solution to a complex texture discrimination problem (more completely discussed in (69)). The way forward may, therefore, entail improving our understanding of what textures demonstrate fixation consistency across expert radiologists. Rather than solely focusing on enhancing search strategies which may be theoretically appealing, but (1) fail to consider the unique characteristics of each visual search task (56), (2) are not individualized, and (3) may not be employed by expert radiologists (31,56), perceptual learning heuristics could be designed to train for enhanced detection of particular textures, resulting in enhanced detection and decreased omission errors.

A more refined approach in our understanding of perceptual expertise, and better models to test perception in radiologists, can also aid in improving training—isolating "at risk" residents and attendings for targeted interventions. As a field dominated by a primarily *perceptual* task, radiology needs a greater understanding of perceptual expertise to improve accuracy, reduce error, and improve patient care.

ACKNOWLEDGMENTS

This work was supported by the NSF Award # 1523614 to SMC and Award # 1734887 to SMC and SLM.

REFERENCES

1. Gray JE, Taylor KW, Hobbs BB. Detection accuracy in chest radiography. AJR Am J Roentgenol 1978; 131(2):247–253.
2. Garland LH. On the scientific evaluation of diagnostic procedures. Radiology 1949; 52(3):309–328.
3. Berlin L.Accuracy of diagnostic procedures: has it improved over the past five decades?AJR Am J Roentgenol. 2007; 188(5):1173-8.
4. Waite S, Scott J, Gale B, et al. Interpretive Error in Radiology. AJR Am J Roentgenol 2017; 208(4):739–749.
5. Siegle RL, Baram EM, Reuter SR, et al. Rates of disagreement in imaging interpretation in a group of community hospitals. Acad Radiol 1998; 5(3):148–154.
6. Bruno MA, Walker EA, Abujudeh HH. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. RadioGraphics 2015; 35(6):1668–1676.
7. Forrest JV, Friedman PJ. Radiologic errors in patients with lung cancer. The Western journal of medicine 1981; 134(6):485–490.
8. Muhm JR, Miller WE, Fontana RS, et al. Lung cancer detected during a screening program using four-month chest radiographs. Radiology 1983; 148(3):609–615.
9. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? Insights into Imaging 2017; 8(1):171–182.
10. Herzog R, Elgort DR, Flanders AE, et al. Variability in diagnostic error rates of 10 MRI centers performing lumbar spine MRI examinations on the same patient within a 3-week period. The spine journal: official journal of the North American Spine Society 2017; 17(4):554–561.
11. Mucci B, Murray H, Downie A, et al. Interrater variation in scoring radiological discrepancies. Br J Radiol 2013; 86(1028):20130245.
12. Bender LC, Linnau KF, Meier EN, et al. Interrater Agreement in the Evaluation of Discrepant Imaging Findings With the Radpeer System. American Journal of Roentgenology 2012; 199(6):1320–1327.
13. Abujudeh HH, Boland GW, Kaewlai R, et al. Abdominal and pelvic computed tomography (CT) interpretation: discrepancy rates among experienced radiologists. Eur Radiol 2010; 20(8):1952–1957.
14. Lauritzen PM, Andersen JG, Stokke MV, et al. Radiologist-initiated double reading of abdominal CT: retrospective analysis of the clinical importance of changes to radiology reports. BMJ Quality & Safety 2016; 25(8):595–603.
15. Kim YW, Mansfield LT. Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors. AJR Am J Roentgenol 2014; 202 (3):465–470.

16. Renfrew DL, Franken Jr. EA, Berbaum KS, et al. Error in radiology: classification and lessons in 182 cases presented at a problem case conference. Radiology 1992; 183(1):145–150.
17. L. Kundel H.How to minimize perceptual error and maximize expertise in medical imaging. 2007; 651508.
18. Funaki B, Szymski G, Rosenblum J. Significant on-call misses by radiology residents interpreting computed tomographic studies: Perception versus cognition. Emergency Radiology 1997; 4(5):290–294.
19. Degnan AJ, Ghobadi EH, Hardy P, et al. Perceptual and Interpretive Error in Diagnostic Radiology-Causes and Potential Solutions. Acad Radiol 2018; 26(6):833–845.
20. Baker SR, Patel RH, Yang L, et al. Malpractice suits in chest radiology: an evaluation of the histories of 8265 radiologists. Journal of thoracic imaging 2013; 28(6):388–391.
21. McCamy MB, Otero-Millan J, Di Stasi LL, et al. Highly informative natural scene regions increase microsaccade production during visual scanning. J Neurosci 2014; 34(8):2956–2966.
22. van der Gijp A, Ravesloot CJ, Jarodzka H, et al. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Advances in Health Sciences Education 2017; 22(3):765–787.
23. Wolfe JM, Evans KK, Drew T, et al. How Do Radiologists Use the Human Search Engine? Radiat Prot Dosimetry 2016; 169(1-4):24–31.
24. Nakashima R, Komori Y, Maeda E, et al. Temporal Characteristics of Radiologists' and Novices' Lesion Detection in Viewing Medical Images Presented Rapidly and Sequentially. Front Psychol 2016; 7:1553.
25. Mallett S, Phillips P, Fanshawe TR, et al. Tracking eye gaze during interpretation of endoluminal three-dimensional CT colonography: visual perception of experienced and inexperienced readers. Radiology 2014; 273(3):783–792.
26. Bertram R, Kaakinen J, Bensch F, et al. Eye Movements of Radiologists Reflect Expertise in CT Study Interpretation: A Potential Tool to Measure Resident Development. Radiology 2016; 281(3):805–815.
27. Puddy E, Hill C. Interpretation of the chest radiograph. Continuing Education in Anaesthesia Critical Care & Pain 2007; 7(3):71–75.
28. Goodman LR, Felson B. Felson's principles of chest roentgenology. Philadelphia, Pa.; Edinburgh: Elsevier Saunders, 2007 48.
29. Kok EM, Jarodzka H, de Bruin AB, et al. Systematic viewing in radiology: seeing more, missing less? Adv Health Sci Educ Theory Pract 2015; 21(1):189–205.
30. van Geel K, Kok EM, Dijkstra J, et al. Teaching Systematic Viewing to Final-Year Medical Students Improves Systematicity but Not Coverage or Detection of Radiologic Abnormalities. J Am Coll Radiol 2017; 14 (2):235–241.
31. Kundel HL, La Follette Jr. PS. Visual search patterns and experience with radiological images. Radiology 1972; 103(3):523–528.
32. Tuddenham WJ, Calvert WP. Visual search patterns in roentgen diagnosis. Radiology 1961; 76:255–256.
33. Venjakob AC, Mello-Thoms CR. Review of prospects and challenges of eye tracking in volumetric imaging. J Med Imaging (Bellingham) 2016; 3 (1):011002.
34. den Boer L, van der Schaaf MF, Vincken KL, et al. Volumetric image interpretation in radiology: scroll behavior and cognitive processes. Advances in Health Sciences Education 2018; 23(4):783–802.
35. van der Gijp A, van der Schaaf MF, van der Schaaf IC, et al. Interpretation of radiological images: towards a framework of knowledge and skills. Advances in Health Sciences Education 2014; 19(4):565–580.
36. Drew T, Vo ML-H, Olwal A, et al. Scanners and drillers: characterizing expert visual search through volumetric images. J Vis 2013; 13(10):1–13.
37. Kelahan LC, Fong A, Blumenthal J, et al. The Radiologist's Gaze: Mapping Three-Dimensional Visual Search in Computed Tomography of the Abdomen and Pelvis. J Digit Imaging 2018; 32(2):234–240.
38. Andriole KP, Wolfe JM, Khorasani R, et al. Optimizing Analysis, Visualization, and Navigation of Large Image Data Sets: One 5000-Section CT Scan Can Ruin Your Whole Day. Radiology 2011; 259(2):346–362.
39. Eckstein MP, Lago MA, Abbey CK. The role of extra-foveal processing in 3D imaging. Proc SPIE Int Soc Opt Eng 2017; 10136:101360E.
40. Miller Jr. WT, Marinari LA, Barbosa Jr. E. et al. Small pulmonary artery defects are not reliable indicators of pulmonary embolism. Annals of the American Thoracic Society 2015; 12(7):1022–1029.
41. Rubin GD, Roos JE, Tall M, et al. Characterizing search, recognition, and decision in the detection of lung nodules on CT scans: elucidation with eye tracking. Radiology 2015; 274(1):276–286.
42. van der Gijp A, Vincken KL, Boscardin C, et al. The Effect of Teaching Search Strategies on Perceptual Performance. Acad Radiol 2017; 24 (6):762–767.
43. Wen G, Aizenman A, Drew T, et al. Computational assessment of visual search strategies in volumetric medical images. J Med Imaging (Bellingham) 2016; 3(1):015501.
44. Kok EM, Jarodzka H, de Bruin AB, et al. Systematic viewing in radiology: seeing more, missing less? Adv Health Sci Educ Theory Pract 2016; 21(1):189–205.
45. Kok EM. Eye tracking: the silver bullet of competency assessment in medical image interpretation? Perspectives on medical education 2019; 8(2):63–64.
46. Aizenman A, Drew T, Ehinger KA, et al. Comparing search patterns in digital breast tomosynthesis and full-field digital mammography: an eye tracking study. Journal of Medical Imaging 2017; 4(4):1–10.
47. Birchall D. Spatial ability in radiologists: a necessary prerequisite? Br J Radiol 2015; 88(1049):20140511.
48. Sunday MA, Donnelly E, Gauthier I. Individual differences in perceptual abilities in medical imaging: the Vanderbilt Chest Radiograph Test. Cogn Res Princ Implic 2017; 2(1):36.
49. Nodine CF, Krupinski, Perceptual EA. skill, radiology expertise, and visual test performance with NINA and WALDO. Academic radiology 1998; 5(9):603–612.
50. Kelly B, Rainford LA, McEntee MF, et al. Influence of radiology expertise on the perception of nonmedical images. J Med Imaging (Bellingham) 2018; 5(3):031402.
51. Smoker WR, Berbaum KS, Luebke NH, et al. Spatial perception testing in diagnostic radiology. AJR Am J Roentgenol 1984; 143(5):1105–1109.
52. Nodine C, Kundel H. The cognitive side of visual search in radiology. Eye Movements from Physiology to Cognition: Elsevier 1987: 573–582.
53. Nodine CF, Kundel HL. Using eye movements to study visual search and to improve tumor detection. Radiographics 1987; 7(6):1241–1250.
54. Sheridan H, Reingold EM. The Holistic Processing Account of Visual Expertise in Medical Image Perception: A Review. Front Psychol 2017; 8:1620.
55. Krupinski EA. The role of perception in imaging: past and future. Semin Nucl Med 2011; 41(6):392–400.
56. Gandomkar Z, Mello-Thoms C. Visual search in breast imaging. The British Journal of Radiology 2019; 92(1102):20190057.
57. Kundel HL, Nodine CF, Conant EF, et al. Holistic component of image perception in mammogram interpretation: gaze-tracking study. Radiology 2007; 242(2):396–402.
58. Carrigan AJ, Wardle SG, Rich AN. Finding cancer in mammograms: if you know it's there, do you know where? Cogn Res Princ Implic 2018; 3(1):10.
59. Kundel HL, Nodine CF. Interpreting chest radiographs without visual search. Radiology 1975; 116(3):527–532.
60. Evans KK, Georgian-Smith D, Tambouret R, et al. The gist of the abnormal: above-chance medical decision making in the blink of an eye. Psychon Bull Rev 2013; 20(6):1170–1175.
61. Oestmann JW, Greene R, Kushner DC, et al. Lung lesions: correlation between viewing time and detection. Radiology 1988; 166(2):451–453.
62. Evans K, Birdwell R, Georgian-Smith D, et al. Discrimination and Localization of Abnormalities in Mammograms from a Global Signal. In: Radiological Society of North America 2010 Scientific Assembly and Annual Meeting, Chicago, IL; 2010.
63. Drew T, Evans K, Vo ML, et al. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? Radiographics 2013; 33(1):263–274.
64. Chin MD, Evans KK, Wolfe JM, et al. Inversion effects in the expert classification of mammograms and faces. Cogn Res Princ Implic 2018; 3:31.
65. Evans KK, Haygood TM, Cooper J, et al. A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. Proc Natl Acad Sci U S A 2016; 113(37):10292–10297.
66. Brady TF, Konkle T, Alvarez GA. A review of visual memory capacity: Beyond individual items and toward structured representations. Journal of vision 2011; 11(5):4.
67. Akyurek EG, Kappelmann N, Volkert M, et al. What You See Is What You Remember: Visual Chunking by Temporal Integration Enhances Working Memory. Journal of cognitive neuroscience 2017; 29 (12):2025–2036.

68. Gobet F, Lane PC, Croker S, et al. Chunking mechanisms in human learning. Trends in cognitive sciences 2001; 5(6):236–243.
69. Waite S, Grigorian A, Alexander RG, et al. Analysis of Perceptual Expertise in Radiology – Current Knowledge and a New Perspective. Frontiers in Human Neuroscience 2019; 13(213):213.
70. Litchfield D, Donovan T. The flash-preview moving window paradigm: Unpacking visual expertise one glimpse at a time. Frontline Learning Research 2017; 5(3):66–80.
71. Kelly BS, Rainford LA, Darcy SP, et al. The Development of Expertise in Radiology: In Chest Radiograph Interpretation, "Expert" Search Pattern May Predate "Expert" Levels of Diagnostic Accuracy for Pneumothorax Identification. Radiology 2016; 280(1):252–260.
72. Kok EM, de Bruin ABH, Robben SGF, et al. Looking in the Same Manner but Seeing it Differently: Bottom-up and Expertise Effects in Radiology. Applied Cognitive Psychology 2012; 26(6):854–862.
73. Ravesloot CJ, van der Schaaf MF, Kruitwagen CLJJ, et al. Predictors of Knowledge and Image Interpretation Skill Development in Radiology Residents. Radiology 2017; 284(3):758–765.
74. Rutgers DR, van Raamt F, Ten Cate TJ. Development of competence in volumetric image interpretation in radiology residents. BMC Med Educ 2019; 19(1):122.
75. Kellman PJ, Garrigan P. Perceptual learning and human expertise. Phys Life Rev 2009; 6(2):53–84.
76. Kundel HL, Nodine CF. A visual concept shapes image perception. Radiology 1983; 146(2):363–368.
77. Chen W, HolcDorf D, McCusker MW, et al. Perceptual training to improve hip fracture identification in conventional radiographs. PLoS One 2017; 12(12):e0189192.
78. Sowden PT, Davies IR, Roling P. Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity? J Exp Psychol Hum Percept Perform 2000; 26 (1):379–390.
79. Adams M, Chen W, Holcdorf D, et al. Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures. Journal of Medical Imaging and Radiation Oncology 2019; 63 (1):27–32.
80. Becker GJ, Bosma JL, Guiberteau MJ, et al. ABR Examinations: The Why, What, and How. Radiology 2013; 268(1):219–227.
81. Alderson PO, Becker GJ. The new requirements and testing for American Board of Radiology certification in diagnostic radiology. Radiology 2008; 248(3):707–709.
82. Kerridge WD, Gunderman RB. The Validity and Timing of the ABR Core Exam. Acad Radiol 2016; 23(9):1176–1179.
83. Hall FM. The ABR and resident recall "cheating". Radiology 2012; 263 (2):323–325.
84. American Board of Radiology. New Question types: Part 1 and 2. https://www.theabr.org/medical-physics/initial-certification/new-questions-parts1-2.
85. Jackson VP, Balfe DM, Guiberteau MJ. Counterpoint: Why Things Are Going Right With the ABR Examinations. Journal of the American College of Radiology 2016; 13(11):1361–1362.
86. Heitkamp DE, Kamer AP, Smith J, et al. Point: What Went Wrong With the ABR Examinations? Journal of the American College of Radiology 2016; 13(11):1359–1360.
87. Berland LL, Berland NW, Berland MW. ABR Psychometric Testing: Analysis of Validity and Effects. J Am Coll Radiol 2018; 15(6):905–910.
88. Guiberteau MJ, Balfe DM, Jackson VP, et al. The ABR Core Exam: What Is the Value? Academic Radiology 2016; 23(9):1180–1182.
89. Castells X, Torá-Rocamora I, Posso M, et al. Risk of Breast Cancer in Women with False-Positive Results according to Mammographic Features. Radiology 2016; 280(2):379–386.
90. American College of Radiology. Case Study: A Simulated experience. https://www.acr.org/-/media/ACR/Files/Case-Studies/IT/A-Simulated-Experience/Imaging3_ASimulatedExperience_February2015.pdf. In: ACR ed.2015; 1-3.
91. Pakdemirli E. Artificial intelligence in radiology: friend or foe? Where are we now and where are we heading? Acta Radiol Open 2019; 8(2):1–5.
92. Botz B.A Coyote That Ran Over the Edge? — Differing Opinions About the Impact of AI on Radiology. Available at: https://chatbotslife.com/a-coyote-that-ran-over-the-edge-differing-opinions-about-the-impact-of-ai-on-radiology-4b5e3860722e. Accessed October 8, 2019.
93. Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians. Proceedings of the National Academy of Sciences 2018; 115(45):11591–11596.
94. Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. PLoS Med 2018; 15(11): e1002697.
95. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine 2019; 25(6):954–961.
96. Langlotz CP, Allen B, Erickson BJ, et al. A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. Radiology 2019; 291(3):781–791.
97. Mirsky Y, Mahler T, Shelef I, et al. CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning. USENIX Security Symposium 2019: 461–478.
98. Liew C. The future of radiology augmented with Artificial Intelligence: A strategy for success. European Journal of Radiology 2018; 102:152–156.
99. Dellermann D, Calma A, Lipusch N, et al. The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. Hawaii International Conference on System Sciences 2019: 1–10.
100. Kasparov G. The chess master and the computer. The New York Review of Books 2010; 57(2):16–19.
101. Tajmir SH, Alkasab TK. Toward Augmented Radiologists: Changes in Radiology Education in the Era of Machine Learning and Artificial Intelligence. Acad Radiol 2018; 25(6):747–750.
102. Duong MT, Rauschecker AM, Rudie JD, et al. Artificial intelligence for precision education in radiology. The British journal of radiology 2019; 92:20190389.
103. Krasne S, Hillman JD, Kellman PJ, et al. Applying perceptual and adaptive learning techniques for teaching introductory histopathology. J Pathol Inform 2013; 4:34.
104. Zhang J, Lo JY, Kuzmiak CM, et al. Using computer-extracted image features for modeling of error-making patterns in detection of mammographic masses among radiology residents. Med Phys 2014; 41 (9):091907.
105. Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Invest Radiol 1978; 13(3):175–181.
106. Krupinski EA. Current perspectives in medical image perception. Atten Percept Psychophys 2010; 72(5):1205–1217.
107. Krupinski EA, Nodine CF, Kundel HL. Enhancing recognition of lesions in radiographic images using perceptual feedback. Optical Engineering 1998; 37(3):813–818.
108. Hu CH, Kundel HL, Nodine CF, et al. Searching for bone fractures: a comparison with pulmonary nodule search. Acad Radiol 1994; 1(1):25–32.
109. Nodine CF, Mello-Thoms C, Kundel HL, et al. Time course of perception and decision making during mammographic interpretation. AJR Am J Roentgenol 2002; 179(4):917–923.
110. Wen G, Rodriguez-Nino B, Pecen FY, et al. Comparative study of computational visual attention models on two-dimensional medical images. J Med Imaging (Bellingham) 2017; 4(2):025503.
111. Waite S, Kolla S, Jeudy J, et al. Tired in the Reading Room: The Influence of Fatigue in Radiology. J Am Coll Radiol 2017; 14(2):191–197.
112. Kagadis GC, Walz-Flannigan A, Krupinski EA, et al. Medical Imaging Displays and Their Use in Image Interpretation. RadioGraphics 2013; 33 (1):275–290.
113. Carmody DP, Kundel HL, Toto LC. Comparison scans while reading chest images. Taught, but not practiced. Invest Radiol 1984; 19(5):462–466.
114. Wolfe JM, Vo ML, Evans KK, et al. Visual search in scenes involves selective and nonselective pathways. Trends Cogn Sci 2011; 15(2):77–84.
115. Pao DG. The lateral femoral notch sign. Radiology 2001; 219(3): 800–801.