

REVIEW ARTICLE

Open Access



A review of eye tracking for understanding and improving diagnostic interpretation

Tad T. Bruny ^{1*}, Trafton Drew², Donald L. Weaver³ and Joann G. Elmore⁴

Abstract

Inspecting digital imaging for primary diagnosis introduces perceptual and cognitive demands for physicians tasked with interpreting visual medical information and arriving at appropriate diagnoses and treatment decisions. The process of medical interpretation and diagnosis involves a complex interplay between visual perception and multiple cognitive processes, including memory retrieval, problem-solving, and decision-making. Eye-tracking technologies are becoming increasingly available in the consumer and research markets and provide novel opportunities to learn more about the interpretive process, including differences between novices and experts, how heuristics and biases shape visual perception and decision-making, and the mechanisms underlying misinterpretation and misdiagnosis. The present review provides an overview of eye-tracking technology, the perceptual and cognitive processes involved in medical interpretation, how eye tracking has been employed to understand medical interpretation and promote medical education and training, and some of the promises and challenges for future applications of this technology.

Keywords: Eye tracking, Medical informatics, Visual perception, Visual search, Medical decision-making

Significance

During patient examinations, image interpretation, and surgical procedures, physicians are constantly accumulating multisensory evidence when inspecting information and ultimately arriving at a diagnostic interpretation. Eye-tracking research has shed light on the dynamics of this interpretive process, including qualitative and quantitative differences that help distinguish and possibly predict successes and errors. This progress affords novel insights into how the interpretive process might be improved and sustained during education, training, and clinical practice. The present review details some of this research and emphasizes future directions that may prove fruitful for scientists, educators, and clinical practitioners interested in accelerating the transition from novice to expert, monitoring and maintaining competencies, developing algorithms to automate error detection and classification, and informing tractable remediation strategies to train the next generation of diagnosticians.

Introduction

Decades of research have demonstrated the involvement of diverse perceptual and cognitive processes during medical image interpretation and diagnosis (Bordage, 1999; Elstein, Shulman, & Sprafka, 1978; Gilhooly, 1990; Kundel & La Follette, 1972; Patel, Arocha, & Zhang, 2005). Broadly speaking, these include visual search and pattern matching, hypothesis generation and testing, and reasoning and problem-solving. As with many more general cognitive tasks, these processes interact dynamically over time via feed-forward and feed-back mechanisms to guide interpretation and decision-making (Brehmer, 1992; Newell, Lagnado, & Shanks, 2015). The reliable involvement of these processes has made them of interest as targets for both clinical research and the design of educational interventions to improve diagnostic decision-making (Crowley, Naus, Stewart, & Friedman, 2003; Custers, 2015; Nabil et al., 2013). Methodologies to investigate mental processes during interpretation and diagnosis have included think-aloud protocols (Lundgr n-Laine & Salanter , 2010), knowledge and memory probes (Gilhooly, 1990; Patel & Groen, 1986), practical exercises (Bligh, Prideaux, & Parsell, 2001; Harden, Sowden, & Dunn, 1984), and tracking physicians' interface navigation

* Correspondence: tbruny01@tufts.edu

¹Center for Applied Brain and Cognitive Sciences, Tufts University, 200 Boston Ave., Suite 3000, Medford, MA 02155, USA

Full list of author information is available at the end of the article

behavior while they inspect visual images (e.g., radiographs, histology slides) (Mercan et al., 2016; Mercan, Shapiro, Brunyé, Weaver, & Elmore, 2017).

Medical researchers have increasingly turned to eye-tracking technology to provide more detailed qualitative and quantitative assessments of how and where the eyes move during interpretation, extending research from other high-stakes domains such as air-traffic control (Martin, Cegarra, & Averty, 2011) and airport luggage screening (McCarley & Carruth, 2004; McCarley, Kramer, Wickens, Vidoni, & Boot, 2004). Studies in the medical domain have provided more nuanced understandings of visual interpretation and diagnostic decision-making in diverse medical specialties including radiology, pathology, pediatrics, surgery, and emergency medicine (Al-Moteri, Symmons, Plummer, & Cooper, 2017; Blondon & Lovis, 2015; van der Gijp et al., 2017). Eye tracking has the potential to revolutionize clinical practice and medical education, with far-reaching implications for the development of automated competency assessments (Bond et al., 2014; Krupinski, Graham, & Weinstein, 2013; Richstone et al., 2010; Tien et al., 2014), advanced clinical tutorials (e.g., watching an expert's eye movements over an image; (Khan et al., 2012; O'Meara et al., 2015)), biologically inspired artificial intelligence to enhance computer-aided diagnosis (Buettner, 2013; Young & Stark, 1963), and the automated detection and mitigation of emergent interpretive errors during the diagnostic process (Ratwani & Trafton, 2011; Tourassi, Mazurowski, Harrawood, & Krupinski, 2010; Voisin, Pinto, Morin-Ducote, Hudson, & Tourassi, 2013).

Eye tracking: technologies and metrics

Modern eye tracking involves an array of infrared or near-infrared light sources and cameras that track the gaze behavior of one (monocular) or both (binocular) eyes (Holmqvist et al., 2011). In most modern systems, an array of non-visible light sources illuminate the eye and produce a corneal reflection (the first Purkinje image); the eye tracker monitors the relationship between this reflection and the center of the pupil to compute vectors that relate eye position to locations in the perceived world (Hansen & Ji, 2010). As the eyes move, the computed point of regard in space also moves. Eye trackers are available in several hardware configurations, including systems with a chin rest for head stabilization, remote systems that can accommodate a limited extent of head movements, and newer mobile eye-wear based systems. Each of these form factors has relative advantages and disadvantages for spatial accuracy (i.e., tracking precision), tracking speed, mobility, portability, and cost (Funke et al., 2016; Holmqvist, Nyström, & Mulvey, 2012). Figure 1 depicts a relatively mobile and contact-free eye-tracking system manufactured by SensoMotoric

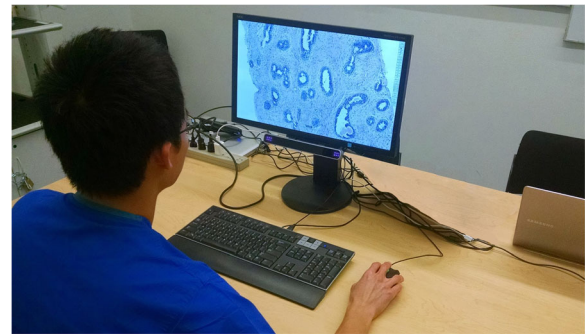


Fig. 1 A remote eye-tracking system (SensoMotoric Instruments' Remote Eye-tracking Device – mobile; SMI REDm) mounted to the bottom of a computer monitor. In this study, a participating pathologist is inspecting a digital breast biopsy (Brunyé, Mercan, et al., 2017)

Instruments (SMI; Berlin, Germany), the Remote Eye-tracking Device – mobile (REDm).

Eye trackers provide several measures of visual behavior that are relevant for understanding the interpretive process; these are categorically referred to as movement measures, position measures, numerosity measures, and latency measures (Holmqvist et al., 2011). Before describing these, it is important to realize that the eye is constantly moving between points of fixation. Fixations are momentary pauses of eye gaze at a spatial location for a minimum amount of time (e.g., >99 ms), and the movements between successive fixations are called saccades (Liversedge & Findlay, 2000). Movement measures quantify the patterns of eye movements through space during saccades, including the distance between successive saccades (degrees of saccade amplitude) and the speed of saccades (typically average or peak velocity). Position measures quantify the location of the gaze in Cartesian coordinate space, such as the coordinate space of a computer monitor, or a real-world scene captured through a forward-view camera. Numerosity measures quantify the frequency with which the eyes fixate and saccade while perceiving a scene, such as how many fixations and saccades have occurred during a given time, and how those counts might vary as a function of position (and the visual information available at different positions). Finally, latency measures allow for an assessment of the temporal dynamics of fixations and saccades, including first and subsequent fixation durations and saccade duration. Table 1 provides an overview of commonly used eye-tracking measures, and current theoretical perspectives on their relationships to perceptual and cognitive processing.

Eye tracking in medical interpretation

Some of the earliest research using eye tracking during medical image interpretation was done during x-ray film inspection (Kundel & Nodine, 1978). In this task, radiologists search chest x-ray films for evidence of lung

Table 1 A taxonomy relating commonly used eye-tracking metrics and their respective units to perceptual and cognitive processes of interest to researchers

Measure	Units	Description
Fixation count	Frequency count	The number of times the eye fixates in a particular region of interest, related to at least: the saliency of the area, the informational value of the area, how much information is available in a single fixation, or the processing difficulty of the information (Findlay & Gilchrist, 2008; Henderson & Hollingworth, 1998; Henderson, Malcolm, & Schandl, 2009)
Regressive fixation count	Frequency count	Re-fixating a previously fixated region, to resolve ambiguity or other processing difficulties (Spivey & Tanenhaus, 1998; Underwood & Radach, 1998)
Fixation duration	Milliseconds	How long the eye fixates on a region prior to a saccade, related to the difficulty in processing the information in that region, the value of information available in that region, the time needed to plan the next saccade, and the predicted value of information available following the next saccade (Findlay & Gilchrist, 2008; Rayner, 1998; Sumner, 2011)
Amplitude	Degrees	The magnitude of a saccade, influenced by how much information can be processed in the area of a single fixation, and the distance to the next planned fixation target (Rayner, 1998)
Saccade peak velocity	Degrees/second	The maximum speed achieved within a saccade, related to physiological arousal, mental workload, or the predicted value of information available at the subsequent fixation (Di Stasi, Catena, Cañas, Macknik, & Martinez-Conde, 2013; Montagnini & Chelazzi, 2005; Xu-Wilson et al., 2009)
Blink rate or inter-blink interval	Frequency count/time or milliseconds	The number of eye blinks detected by an eye tracker's algorithms, inversely related to physiological arousal, wakefulness, processing difficulty, motivation, and mental workload (Holmqvist et al., 2011; Siegle, Ichikawa, & Steinhauer, 2008)
Blink amplitude and blink duration	Milliseconds	The extent and duration of an eye blink (temporary closure) event, inversely related to physiological arousal, wakefulness, processing difficulty, motivation, and mental workload (Holmqvist et al., 2011; Ingre, Åkerstedt, Peters, Anund, & Kecklund, 2006).
Phasic pupil diameter	Millimeter diameter	Rapid and dramatic pupil diameter changes related to processing task- and goal-relevant information, and exploiting that information to perform a task (Beatty, 1982; Laeng, Sirois, & Gredeback, 2012)
Tonic pupil diameter	Millimeter diameter	Sustained pupil diameter changes that establish a new baseline diameter from which phasic responses deviate, related to sustained cognitive processing, task difficulty, cognitive effort, arousal, and vigilance (Laeng et al., 2012; Siegle et al., 2008).

nodules; Kundel and Nodine were interested in whether radiologists were making errors of visual search versus errors of recognition and/or decision-making. A search error would be evidenced by a failure to fixate on a nodule, and a recognition or decision error would occur when a fixation on a nodule is not followed by a successful identification and diagnosis. To further differentiate errors of recognition versus decision-making, Kundel and Nodine distinguished trials where the radiologist fixated within 2.8° of a nodule for greater than or less than 600 ms. If the fixation occurred for less than 600 ms this was considered a recognition error, and if greater than 600 ms it was considered a decision error. The former was considered a failure to disembed the nodule from the background noise (despite fixating on it), and the latter was considered a successful recognition of a nodule without appropriately mapping it to diagnostic criteria. Their results demonstrated that about 30% of all errors were due to a failed search. About 25% of errors were due to a recognition failure, and the remaining 45% of errors were due to decision failure. Thus, interpretive errors were primarily driven by failures of recognition and decision-making, rather than failures of search (Kundel & Nodine, 1978). In other words, radiologists would fixate upon and process the critical visual

information in a scene but fail to successfully map that information to known schemas and/or candidate diagnoses. A follow-up study confirmed that fixations over 300 ms did not improve recognition, but did improve decision accuracy; furthermore, fixations within 2° of the nodule were associated with higher recognition accuracy (Carmody, Nodine, & Kundel, 1980). These early studies suggest that eye tracking can be a valuable tool for helping dissociate putative sources of error during medical image interpretation (i.e., search, recognition, and decision-making), given that high-resolution foveal vision appears to be critical for diagnostic interpretation.

Over the past four decades since this original research, eye tracking has been expanded to understanding diagnostic interpretation in several medical specializations, including radiology, breast pathology, general surgery, neurology, emergency medicine, anesthesiology, ophthalmology, and cardiology (Balslev et al., 2012; Berbaum et al., 2001; Brunyé et al., 2014; Giovinco et al., 2015; Henneman et al., 2008; Jungk, Thull, Hoefl, & Rau, 2000; Krupinski et al., 2006; Kundel, Nodine, Krupinski, & Mello-Thoms, 2008; Matsumoto et al., 2011; O'Neill et al., 2011; Sibbald, de Bruin, Yu, & van Merriënboer, 2015; Wood, Batt, Appelboam, Harris, & Wilson, 2014). In general, these eye-tracking studies have found evidence of

reliable distinctions between three types of error-making in diagnostic interpretation: search errors, recognition errors, and decision errors. Each of these error types carries implications for diagnostic accuracy and, ultimately, patient quality of life and well-being. We review each of these in turn, below.

Search errors

A search error occurs when the eyes fail to fixate a critical region of a visual scene, rendering a feature undetected; these have also been labeled as scanning errors because the critical feature was not in the scan path (Cain, Adamo, & Mitroff, 2013). For example, a radiologist failing to fixate a lung nodule (Manning, Ethell, Donovan, & Crawford, 2006), a pathologist failing to fixate large nucleoli in pleomorphic cells (Brunyé, Mercan, Weaver, & Elmore, 2017), or a neuro-radiologist failing to fixate a cerebral infarction (Matsumoto et al., 2011). Theoretically, if the diagnostician has not fixated a diagnostically relevant region of a medical image then successful search has not occurred, and without it, recognition and decision-making are not possible.

Several perceptual and cognitive mechanisms have been proposed to account for why search errors occur, including low target prevalence, satisfaction of search, distraction, and resource depletion. *Low target prevalence* refers to a situation when a diagnostic feature is especially rare. For example, a malignant tumor appearing in a screening mammography examination has a very low prevalence rate at or below 1% of all cases reviewed (Gur et al., 2004). Low prevalence is associated with higher rates of search failure; previous research has shown that when target prevalence was decreased from 50 to 1%, detection rates fell from approximately 93 to 70%, respectively (Wolfe, Horowitz, & Kenner, 2005). Although much of the research on the low prevalence effect has focused on basic findings with naïve subjects, research has also shown that low prevalence also influences diagnostic accuracy in a medical setting (Eggin & Feinstein, 1996; Evans, Birdwell, & Wolfe, 2013). Most notably, Evans and colleagues compared performance under typical laboratory conditions, where target prevalence is high (50% of cases), and when the same cases were inserted into regular workflow, where target prevalence is low (<1% of cases) they found that false-negative rates were substantially elevated at low target prevalence (Evans et al., 2013). As a diagnostician searches a medical image, they must make a decision of when to terminate a search (Chun & Wolfe, 1996; Hong, 2005). In the case of low target prevalence, search termination is more likely to occur prior to detecting a target (Wolfe & Van Wert, 2010).

How exactly a search termination decision emerges during a diagnostician's visual search process is unknown,

though it is likely that there are multiple smaller decisions occurring during the search process: as the diagnostician detects individual targets in the medical image, they must decide whether it is the most diagnostically valuable target (and thus terminate search), or whether they believe there is a rare but more valuable target that might be found with continued search (Rich et al., 2008). The risk is that after finding a single target a diagnostician may terminate search prematurely and fail to detect a target with higher value for a correct diagnosis. This phenomenon was originally coined *satisfaction of search*, when radiologists would become satisfied with their interpretation of a medical image after identifying one lesion, at the expense of identifying a second more important lesion (Berbaum et al., 1990; Smith, 1967). These sorts of errors may be a consequence of Bayesian reasoning based on prior experience: the diagnostician may not deem additional search time justifiable for a target that is exceedingly unlikely to be found (Cain, Vul, Clark, & Mitroff, 2012). More recently, Berbaum and colleagues demonstrated that satisfaction of search alone may not adequately describe the search process (Berbaum et al., 2015; Krupinski, Berbaum, Schartz, Caldwell, & Madsen, 2017). Specifically, detecting a lung nodule on a radiograph did not adversely affect the subsequent detection of additional lung nodules; however, it did alter observers' willingness to report the detected nodules. The authors suggest that detecting a target during search may not induce search termination, but rather change response thresholds during a multiple-target search.

Once a diagnostician finds one target, there is no guarantee that it is the critical feature that will assist in rendering an appropriate diagnosis. It is often the case that critical features are passed over because they are not only low prevalence but also low salience; in other words, they might not stand out visually (in terms of their brightness, contrast, or geometry (Itti & Koch, 2000)) relative to background noise. Research with neurologists and pathologists has demonstrated that novice diagnosticians, such as medical residents, tend to detect features with high visual salience sooner and more often than experienced diagnosticians; this focus on highly salient visual features can be at the cost of neglecting the detection of critical features with relatively low visual salience (Brunyé et al., 2014; Matsumoto et al., 2011). In one study, not only did novice pathologists tend to fixate more on visually salient but diagnostically irrelevant regions, they also tended to re-visit those regions nearly three times as often as expert pathologists (Brunyé et al., 2014). As diagnosticians gain experience with a diverse range of medical images, features, and diagnoses, they develop more refined search strategies and richer knowledge

that accurately guide visual attention toward diagnostically relevant image regions and away from irrelevant regions, as early as the initial holistic inspection of an image (Kundel et al., 2008). As described in Kundel and colleagues' model, expert diagnosticians are likely to detect cancer on a mammogram before any visual scanning (search) takes place, referred to as an initial holistic, gestalt-like perception of a medical image (Kundel et al., 2008). This discovery led these authors to reconceptualize the expert diagnostic process as involving an initial recognition of a feature, followed by a search and diagnosis (Kundel & Nodine, 2010); this is in contrast to traditional conceptualizations suggesting that search always preceded recognition (Kundel & Nodine, 1978). Unlike experts, during the initial viewing of a medical image novices are more likely to be distracted by highly salient image features that are not necessary for diagnostic interpretation. The extent to which a medical image contains visually salient features that are irrelevant for accurate interpretation may make it more likely a novice pathologist or neurologist will be distracted by those features and possibly fail to detect critical but lower-salience image features. This might be especially the case when high-contrast histology stains or imaging techniques render diagnostically irrelevant (e.g., scar tissue) regions highly salient. Eye tracking is a critical tool for recognizing and quantifying attention toward distracting image regions and has been instrumental in identifying this source of search failure among relatively novice diagnosticians.

In a recent taxonomy of visual search errors, Cain and colleagues demonstrated that working memory resources are an important source of errors (Cain et al., 2013). Specifically, when an observer is searching for multiple features (targets), if they identify one feature they may maintain that feature in working memory while searching for another feature. This active maintenance of previously detected features may deplete working memory resources that could otherwise be used to search for lower-salience and prevalence targets. This is evidenced by high numbers of re-fixations in previously detected regions, suggesting an active "refreshing" of the contents of working memory to help maintain item memory (Cain & Mitroff, 2013). This proposal has not been examined with diagnosticians inspecting medical images, though it suggests that physicians with higher working memory capacity may show higher performance when searching for multiple features, offering an interesting avenue for future research. Together, resource depletion, low target prevalence, satisfaction of search, and distraction may account for search errors occurring across a range of disciplines involving medical image interpretation.

Recognition errors

Eye tracking has been instrumental in demonstrating that fewer than half of interpretive errors are attributed to failed search, suggesting that most interpretive errors arise during recognition and decision-making (Al-Moteri et al., 2017; Carmody et al., 1980; Nodine & Kundel, 1987; Samuel, Kundel, Nodine, & Toto, 1995). Recognition errors occur when the eyes fixate a feature, but the feature is not recognized correctly or not recognized as relevant or valuable for the search task. Recognition is an example of attentional mechanisms working together to dynamically guide attention toward features that may be of diagnostic relevance and mapping them to stored knowledge. One way of parsing eye movements into successful versus failed recognition of diagnostically relevant features is to assess fixation durations on critical image regions (Kundel & Nodine, 1978; Mello-Thoms et al., 2005). In this method, individual fixation durations are parsed into two categories using a quantitative threshold. For example, Kundel and Nodine used a 600-ms threshold, and Mello-Thoms and colleagues used a 1000-ms threshold; fixation durations shorter than the threshold indicated failed recognition, whereas durations lengthier than the threshold indicated successful recognition (Kundel & Nodine, 1978; Mello-Thoms et al., 2005). Thus, if a feature (e.g., a lung nodule) was fixated there was successful search, and if it was fixated for longer than the threshold there was successful recognition. Under the assumption that increased fixation durations indicate successful recognition, if a participant fixates on a particular region for longer than a given threshold then any subsequent diagnostic error must be due to failed decision-making.

Using fixation durations to identify successful recognition is an imperfect approach; it is important to note that lengthier fixation durations are also associated with difficulty disambiguating potential interpretations of a feature (Brunyé & Gardony, 2017). In other words, while previous research assumes that lengthy fixation durations indicate successful recognition, they can also indicate the perceptual uncertainty preceding incorrect recognition. This is because a strategic shift of attention toward a particular feature is evident in oculomotor processes, for instance with longer fixations, regardless of whether recognition has proceeded accurately (Heekeren, Marrett, & Ungerleider, 2008). Thus, one can only be truly certain that successful recognition has occurred (i.e., mapping a perceived feature to an accurate knowledge structure) if converging evidence is gathered during the interpretive process.

Consistent with this line of thinking, Manning and colleagues found that false-positives when examining chest radiographs were typically associated with longer cumulative

dwell time than true-positives (Manning et al., 2006). Other methods such as think-aloud protocols and feature annotation may prove especially valuable to complement eye tracking in these situations: when a diagnostician recognizes a feature, they either say it aloud (e.g., “I see cell proliferation”) or annotate the feature with a text input (Pinnock, Young, Spence, & Henning, 2015). These explicit feature recognitions can then be assessed for their accuracy and predictive value toward accurate diagnosis.

In addition to measuring the ballistic movements of the eyes, eye trackers also provide continuous recordings of pupil diameter. Pupil diameter can be valuable for interpreting cognitive states and can be used to elucidate mental processes occurring during medical image interpretation. Pupil diameter is constantly changing as a function of both contextual lighting conditions and internal cognitive states. Alterations of pupil diameter reflecting cognitive state changes are thought to reflect modulation of the locus coeruleus-norepinephrine (LC-NE) system, which indexes shifts from exploration to exploitation states (Aston-Jones & Cohen, 2005; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010). Specifically, when the brain interprets a bottom-up signal (e.g., a salient region that attracts an initial fixation) as highly relevant to a task goal, it will send a top-down signal to selectively orient attention to that region. When that occurs, there is a transient increase in pupil diameter that is thought to reflect a shift from exploring the scene (i.e., searching) to exploiting perceived information that is relevant to the task (Privitera, Renninger, Carney, Klein, & Aguilar, 2010; Usher, Cohen, Servan-Schieber, Rajkowski, & Aston-Jones, 1999). Recent research has demonstrated that during fixation on a scene feature, the time-course of pupil diameter changes can reveal information about an observer’s confidence in their recognition of the feature (Brunyé & Gardony, 2017). Specifically, features that are highly difficult to resolve and recognize cause a rapid pupil dilation response within a second of fixation on the feature. This opens an exciting avenue for using converging evidence, perhaps from fixation duration, pupil diameter, and think-aloud protocols, to more effectively disentangle the instances when lengthy fixations on image features are associated with successful or unsuccessful recognition. In the future, algorithms that can automatically detect instances of successful or failed recognition during fixation may prove particularly valuable for enabling computer-based feedback for trainees.

Decision errors

As observers gather information about a scene, including searching and recognizing features as relevant to task goals, they begin to formulate hypotheses regarding

candidate diagnoses. In some cases, a hypothesis may exist prior to visual inspection of an image (Ledley & Lusted, 1959). The main function of examining a visual image and recognizing features is to develop and test diagnostic hypotheses (Sox, Blatt, Higgins, & Marton, 1988). Developing and testing hypotheses is a cyclical process that involves identifying features that allow the observer to select a set of candidate hypotheses, gathering data to test each hypothesis, and confirming or disconfirming a hypothesis. If the clinician has confirmed a hypothesis, the search may terminate; search may continue if the clinician identifies potential support for multiple hypotheses (e.g., diagnoses with overlapping features) and must continue in the service of differential diagnosis. If the clinician has disconfirmed one of several hypotheses but has not confirmed a single hypothesis, the cyclical process continues; the process also continues under conditions of uncertainty when no given hypotheses have been ruled in or out (Kassirer, Kopelman, & Wong, 1991). It is also important to keep in mind that several diagnoses fall on a spectrum with categorical delineations, with the goal of identifying the highest diagnostic category present in a given image. For instance, a breast pathologist examining histological features may categorize a case as benign, atypia, ductal (DCIS) or lobular carcinoma in situ, or invasive carcinoma (Lester & Hicks, 2016). Given that the most advanced diagnosis is the most important for prognosis and treatment, even if a less advanced hypothesis is supported (e.g., atypia), the pathologist will also spend time ruling out the more advanced diagnoses (e.g., carcinoma in situ, invasive). This may be especially the case when diagnostic features can only be perceived at high-power magnification levels, rendering the remainder of the image immediately imperceptible and making it necessary to zoom out to consider other regions.

In an ideal scenario, critical diagnostic features are detected during search and recognized, which leads the clinician to successfully develop and test hypotheses and produce an accurate diagnosis. In the real world, errors emerge at every step of that process. While decision-related errors may not be readily detected in existing eye-tracking metrics, some recent research suggests that relatively disorganized movements of the eyes over a visual image may indicate higher workload, decision uncertainty, and a higher likelihood of errors (Brunyé, Haga, Houck, & Taylor, 2017; Fabio et al., 2015). Specifically, tracking the entropy of eye movements can indicate relatively disordered search processes that do not follow a systematic pattern. In this case, entropy is conceptualized as the degree of energy dispersal of eye fixations across the screen in a relatively random pattern. Higher fixation entropy might indicate relative uncertainty in the diagnostic decision-making process.

Furthermore, tonic pupil diameter increases can indicate a higher mental workload involved in a decision-making task (Mandrick, Peysakhovich, Rémy, Lepron, & Causse, 2016). No studies have examined the entropy of eye movements during medical image interpretation, and to our knowledge only one has examined pupil diameter (Mello-Thoms et al., 2005), revealing an exciting avenue for continuing research. Specifically, continuing research may find value in combining fixation entropy and pupil diameter to identify scenarios in which successful lesion detection and recognition has occurred, but the clinician is having difficulty arriving at an appropriate decision.

Implications for medical education

Eye tracking may provide innovative opportunities for medical education, training, and competency assessment (Ashraf et al., 2018). Most existing research in this regard leverages the well-established finding that experts move their eyes differently from novices (Brunyé et al., 2014; Gegenfurtner, Lehtinen, & Säljö, 2011; Krupinski, 2005; Krupinski et al., 2006; Kundel et al., 2008; Lesgold et al., 1988). Thus, the premise is that educators can use eye tracking to demonstrate, train, and assess gaze patterns during medical education, possibly accelerating the transition from novice to expert.

Competency-based medical education (CBME) is intended to produce health professionals who consistently demonstrate expertise in both practice and certification (Aggarwal & Darzi, 2006). Though the concept of CBME has been around for several decades, formal frameworks for competency training and assessment have been more recently developed by CanMEDS, the Outcome Project of the US Accreditation Council for Graduate Medical Education (ACGME), and the Scottish Doctor (Frank & Danoff, 2007; Nasca, Philibert, Brigham, & Flynn, 2012; Simpson et al., 2002; Swing, 2007). In each of these cases, methods were evaluated and implemented for integrating CBME, including new standards for curriculum, teaching, and assessment. Many programs, however, have struggled to create meaningful, relevant, and repeatable outcome-based assessments for use in graduate medical education, residency, and fellowships (Holmboe, Edgar, & Hamstra, 2016).

Eye tracking in medical education

As students develop proficiency in interpreting visual images, they demonstrate refined eye movements that move more quickly and consistently toward diagnostic regions of interest (Richstone et al., 2010). In other words, their eye movements increasingly resemble those of experts as they progress through training. One possible method for facilitating this progression is by showing

students video-based playbacks of expert eye movements, a method called eye-movement modeling examples (EMMEs (Jarodzka et al., 2012)). Eye-movement modeling examples typically involve not only showing a video of expert eye movements, but also the expert's audio narrative of the interpretive process (Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, 2013; van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). The idea that EMMEs can assist education leverages a finding from cognitive neuroscience demonstrating that observing another's actions causes the brain to simulate making that same action (i.e., the brain's "mirror system"), and helps students integrate the new action into their own repertoire (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005; Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006). EMMEs also ground a student's education in concrete examples, provide students with unique expert insights that might otherwise be inaccessible, and help students learn explicit strategies for processing the visual image (Jarodzka et al., 2012).

Outside of the medical domain, EMMEs have been demonstrated to help novice aircraft inspectors detect more faults during search (Sadasivan, Greenstein, Gramopadhye, & Duchowski, 2005), circuitry board inspectors detect more faults during search (Nalanagula, Greenstein, & Gramopadhye, 2006), programmers debug software faster (Stein & Brennan, 2004), students become better readers (Mason, Pluchino, & Tornatora, 2015), and novices solve puzzles faster (Velichkovsky, 1995). In medical domains involving visual image inspection, the viewed action is the sequence of an expert clinician's fixations and saccades over the medical image, along with their verbal narration. Few studies have examined the impact of EMMEs in medical learning; note that we differentiate education from training in this context, with education involving the passive viewing of expert eye movements outside of an immediate training context (i.e., not during active practice). In the first study of this kind, novice radiographers viewed either novice or expert eye movements prior to making a diagnostic interpretation of a chest x-ray (Litchfield, Ball, Donovan, Manning, & Crawford, 2010). Viewing expert *or* novice eye movements improved a novice's ability to locate pulmonary nodules relative to a free search, as long as the depicted eye movements showed a successful nodule search. This result suggests that novices can indeed leverage another's eye movements to more effectively guide their own search behavior. More recently, medical students were shown case videos of infant epilepsy, in one of three conditions (Jarodzka et al., 2012). In the control condition, there was expert narration during video playback. Two experimental conditions displayed the narrated video with overlaid expert eye movements; in one condition, the eye movements were indicated by

a small circle, and in the other condition, there was a “spotlight” around the circle that blurred image regions that were outside of the expert’s focus. Results demonstrated increased diagnostic performance of students after viewing the spotlight condition, suggesting that this specific condition was most effective at conveying expert visual search patterns. Thus, some research suggests that passively viewing an expert’s eye gaze can be advantageous to medical education.

While previewing an expert’s eye movements can facilitate interpretive performance on the same or very similar cases, it is unclear whether EMMEs are supporting strategy development that will transfer to dissimilar cases. Transfer describes the ability to apply knowledge, skills and abilities to novel contexts and tasks that have not been previously experienced (Bransford, Brown, & Cocking, 2000). Transfer can be relatively near-transfer versus far-transfer (Barnett & Ceci, 2002), and is considered a critical trademark of successful learning (Simon, 1983). An example of near-transfer might be a pathologist learning the features and rules for diagnosing DCIS on one case or from text-book examples, and transferring that knowledge and skill to a biopsy with similar features that clearly indicate DCIS (Roads, Xu, Robinson, & Tanaka, 2018). An example of relatively far-transfer would be successfully applying knowledge and skill to a novel biopsy with a unique cellular architecture and challenging features that are less clearly indicative of DCIS and are perhaps borderline between atypical ductal hyperplasia (ADH) and DCIS. More research is needed to understand whether EMMEs promote only near-transfer, or whether multiple EMME experiences can promote relatively far-transfer by promoting perceptual differentiation of features, accurate feature recognition, and more accurate and efficient mapping of features to candidate diagnoses. In other words, can EMMEs move beyond providing explicit hints and cues that enable interpretation and diagnosis in highly similar contexts and cases, to accelerating rule and strategy learning that enhances performance on highly dissimilar contexts and cases (Ball & Litchfield, 2017)? Second, it is worth pointing out that some research has suggested that people may intentionally alter their patterns of eye movements if they know that their eye movements are being monitored or that videos of their eye movements will be replayed to others (Neider, Chen, Dickinson, Brennan, & Zelinsky, 2010; Velichkovsky, 1995). While any such effects appear to be both rare and subtle, they do present a challenge to interpreting whether the effects of EMMEs are at least partially due to the intent of the expert viewer as opposed to being a natural representation of their viewing patterns in normal clinical practice (Ball & Litchfield, 2017).

Eye tracking in medical training

As opposed to a novice passively viewing expert eye-gaze behavior, some studies have examined eye gaze as a training tool. As noted previously, we distinguish education from training by noting that training involves active practice of knowledge and skills, with or without feedback (Kern, Thomas, & Hughes, 1998). In most research to date, eye gaze has been used to provide immediate feedback and guidance for a novice during the active exploration of a visual stimulus. This research leverages several phenomena from the cognitive and instructional sciences. First, cueing attention toward relevant features during a training activity can promote more selective attention to cued areas and help observers remember the cued information and allocate less mental energy to the non-cued areas (De Koning, Tabbers, Rikers, & Paas, 2009). For instance, subtle visual cues, such as a momentary flash of light in a specific scene region, can selectively orient attention to that region for further inspection (Danziger, Kingstone, & Snyder, 1998). Second, watching expert eye movements can help observers recognize and learn organizational strategies for viewing and interpreting visual images, understand the expert’s intent, identify the organizational structure of the images, and better organize perceived information into mental schemas (Becchio, Sartori, Bulgheroni, & Castiello, 2008; Jarodzka et al., 2013; Lobmaier, Fischer, & Schwaninger, 2006). For instance, because experts tend to move their eyes and navigate visual images differently than novices, viewing expert eye movements and patterns of navigation behavior may help observers develop more efficient search strategies. Third, well-organized expert eye movements can help an observer recognize relations within and between images, helping them discriminate similar features and possibly promote transfer to novel cases (Kieras & Bovair, 1984). For instance, an expert may saccade intentionally between features that help the observer effectively discriminate them, possibly helping them form a more thorough understanding of how to distinguish features and associated diagnoses. It is unknown whether this refined knowledge would subsequently enable successful transfer to cases with structures and features at least partially overlapping with the learned case, suggesting an avenue for future research.

One popular way to conceptualize the utility of cueing attention toward relevant scene regions is the *Theory of Hints* (Kirsh, 2009). In this theory, when people attempt to solve problems in the real world, they rely not only upon existing knowledge (including heuristics and biases) but also the effective use of any available mental aids offered by the context. In addition to explicit verbal guidance from an instructor, or explicit feedback on worked examples, hints can also come in the form of

another's eye movements (Ball & Litchfield, 2017), which can implicitly (i.e., subconsciously) or explicitly orient attention and provide information to an observer (Thomas & Lleras, 2009a, b). As evidence for relatively implicit attention guidance, novice lung x-ray interpretation can improve when they receive implicit cueing based on an expert's eye movements (Ball & Litchfield, 2017). In accordance with the Theory of Hints, this guidance likely provided not only a cue to orient attention toward a particular scene region, but also increased the likelihood that the area would be considered in their diagnostic interpretation. Specifically, expert cueing can help a novice calibrate the relevance and importance of a region (Litchfield et al., 2010), which can be complemented by an expert's verbal narration. Thus, it seems that cueing an observer with expert eye movements and narration not only guides attention but can also help the student assess the expert's intentionality and incorporate that information into their emergent interpretation. As additional evidence of this phenomenon, when expert eye gaze is superimposed during a simulated laparoscopic surgery task, novices are not only faster to locate critical diagnostic regions, but also more likely to incorporate that region into their diagnosis and ultimately reduce errors (Chetwood et al., 2012). Similarly, when novice trainees have expert eye gaze during a simulated robotic surgical task, they tended to be faster and more productive in identifying suspicious nodules (Leff et al., 2015). In both cases, cueing a trainee with expert eye movements not only gets them to fixate in a desired region, but also seems to help them understand expert intent, behave more like an expert, and develop a more accurate diagnostic interpretation.

Eye tracking in competency assessment

In addition to cueing attention during image interpretation, eye tracking can also be used as a feedback mechanism following case interpretation. As we noted above, medical training frequently involves explicit feedback by instructors on exams and worked examples. But there are few methods for providing feedback regarding the dynamic interpretive process; for instance, how a microscope was panned and zoomed, which features were inspected, and precisely where in the process difficulties may have arisen (Bok et al., 2013; 2016; Kogan, Conforti, Bernabeo, Iobst, & Holmboe, 2011; Wald, Davis, Reis, Monroe, & Borkan, 2009). Identifying concrete metrics for use in competency assessment is critical for understanding and guiding professional development from novices to experts (Dreyfus & Dreyfus, 1986; Green et al., 2009). Indeed, a "lack of effective assessment methods and tools" is noted as a primary challenge for implementing the Milestones initiative in internal medicine education (Holmboe, Call, & Ficalora,

2016; Holmboe, Edgar, & Hamstra, 2016). The Milestones initiative is intended to provide concrete educational milestones for use in assessment of medical competencies during graduate and post-graduate medical education (Swing et al., 2013). The earliest research examining eye tracking for feedback in medicine leveraged the concept of perceptual feedback, which involves showing an observer the regions they tended to focus on during an image interpretation (Kundel, Nodine, & Krupinski, 1990). This procedure was shown to improve decision-making by providing a clinician with a second opportunity to review suspicious image regions and revise their diagnosis; this procedure might be especially advantageous given that most people do not remember where they looked during a search (Võ, Aizenman, & Wolfe, 2016).

Leveraging the concept of using one's own eye movements as a feedback tool, one recent study suggests that eye tracking may be especially valuable for clinical feedback with emergency medicine residents (Szulewski et al., 2018). In that study, eye movements were tracked in emergency medicine residents during objective structured clinical examinations in a simulation environment. During a subsequent faculty debriefing, residents were led through an individualized debrief that included a review of their eye movements during the clinical examination, with reference to scene features focused on their associated decision-making processes. Results demonstrated that all residents deemed the inclusion of eye tracking in the debriefing as a valuable feedback tool for learning, making them more likely to actively reflect on their learning experience, constructively critique themselves and compare themselves to experts, and plan responses for future clinical scenarios (Szulewski et al., 2018). Thus, eye tracking appears to be a valuable tool for augmenting qualitative feedback of trainee performance with concrete examples and guidance to help them attend to appropriate features and incorporate them into diagnoses.

Future research directions

As eye trackers become increasingly available to consumers, lower cost, portable, and easier to use, research on principled methods for using eye tracking for competency assessment is expected to increase (Al-Moteri et al., 2017). It is worth noting that eye trackers with high temporal and spatial resolution and coverage range (e.g., across large or multiple displays) can still be quite cost prohibitive. As eye trackers develop more widespread use, however, one can readily envision both automated and instructor-guided feedback techniques to help quantify competency and provide grounded examples for individualized feedback. In mammography, recent research demonstrates that tracking eye movements and

using machine-learning techniques can predict most diagnostic errors prior to their occurrence, making it possible to automatically provide cueing or feedback to trainees during image inspection (Voisin et al., 2013). In diagnostic pathology, automated feedback may be possible by parsing medical images into diagnostically relevant versus irrelevant regions of interest (ROIs) using expert annotations and/or automated machine-vision techniques (Brunyé et al., 2014; Mercan et al., 2016; Nagarkar et al., 2016). Once these ROIs are established and known to the eye-tracking system, fixations can be parsed as falling within or outside of ROIs. This method could be used to understand the spatial allocation of attention over a digital image (e.g., a radiograph, histology slide, angiography), and the time-course of that allocation.

While eye tracking provides valuable insights into the distribution of visual attention over a scene, it is important to realize that eye trackers are restricted to monitoring foveal vision. The fovea is a small region in the center of the retina that processes light from the center of the visual field, with a dense concentration of cone receptors that provide high visual acuity (Holmqvist et al., 2011). One popular theoretical assumption is that eye and head movements strategically position the retina to a more advantageous state for gathering information, such as moving your head and eyes toward the source of a sound to reveal its nature and relevance (Xu-Wilson, Zee, & Shadmehr, 2009). Thus, some of what we consider overt visual attention should theoretically be captured by tracking eye movements. On the other hand, it is also well-established that visual attention can be shifted and sustained covertly, allowing one to fixate the eyes on an ostensibly uninteresting or irrelevant feature while covertly attending to another (Liversedge & Findlay, 2000; Treisman & Gelade, 1980). Thus, it remains possible that some of a diagnostician's interpretive process may occur through peripheral vision (parafoveal vision), limiting our interpretation of eye-tracking patterns made during medical image inspection.

Eye trackers are designed to track eye gaze as a series of fixations and saccades; in other words, they are designed to track foveal attention. This means that they are quite good at tracking overt central visual attention, but they are not intended for tracking covert peripheral visual attention (Holmqvist et al., 2011). However, we also know that visual attention can be covertly shifted to other areas of a visual scene without a subsequent overt fixation on that region (Liversedge & Findlay, 2000; Treisman & Gelade, 1980). This is typically considered a major downfall of eye tracking: that many real-world visual tasks likely involve both covert *and* overt visual attention, though eye tracking can only measure the latter. However, more recent research has demonstrated

that microsaccades reflect shifts in covert attention (Meyberg, Werkle-Bergner, Sommer, & Dimigen, 2015; Yuval-Greenberg, Merriam, & Heeger, 2014). Microsaccades are very small saccades that are less than 1° of visual arc and occur very frequently during fixations (about two to three times per second) (Martinez-Conde, Otero-Millan, & MacKnik, 2013). These microsaccades tend to be directional, for instance moving slightly to the left or right of a current fixation point; research has recently demonstrated that these slight directional movements of the eye indicate the orientation of covert attention (Yuval-Greenberg et al., 2014). For example, if you are staring at a point on a screen but monitoring an upper-right area of the periphery for a change, then microsaccades are likely to show a directional shift toward the upper right. Microsaccades are likely to serve many purposes, such as preparing the eye for a subsequent saccade to a peripheral region (Juan, Shorter-Jacobi, & Schall, 2004), but can also provide meaningful metrics of covert attention. With a clinician, it is possible that while they fixated on a given number of regions they also considered additional image regions for fixation (but never visited them). In other words, microsaccades may provide more fine-grained understanding of the strategic search process within individual fixations and allow a more nuanced understanding of which regions might have been ruled-out or ruled-in for subsequent inspection.

Eye tracking also carries value for understanding longitudinal aspects of competency progression in medical education. While diagnostic performance is routinely evaluated through credentialing and certification, we have very little insight into the underlying interpretive process or the process of skills development over time. For instance, within the domain of diagnostic pathology, we know of only one study that examined longitudinal changes in pathology residents' visual expertise (Krupinski et al., 2013). Unfortunately, this prior study is limited by its size and breadth (four residents at a single training location), the restriction of observers' ability to zoom or pan the medical image, and a reliance on the same experimental images each year. Thus, most of our understanding of how image interpretation and diagnostic accuracy and efficiency emerge during professional development is restricted to insights from cross-sectional designs. But we also know that expertise development of medical students and post-graduate resident trainees is a long-term, continuous, and non-linear process. Eye tracking provides an innovative opportunity to enable a large-scale examination of how interpretive and diagnostic skills develop through multi-year residencies and into professional practice. Our current research is examining this exciting possibility.

We have focused primarily on competency development through education and training, and performance differences between novices and experts. However, it is worth pointing out that each individual student and clinician brings a unique set of individual differences to clinical diagnostics that undoubtedly influences the processes of visual search and decision-making. Individual differences include variables such as personality traits and cognitive abilities, and a substantial body of research demonstrates that these variables constantly influence real-world behavior (Motowildo, Borman, & Schmit, 1997). For instance, recent research has demonstrated that experienced radiologists show superior perceptual abilities to novices, as measured with the Vanderbilt Chest Radiograph Test (Sunday, Donnelly, & Gauthier, 2017). Here we consider one individual difference that warrants more consideration in the domains of medical image interpretation and decision-making: working-memory capacity. Generally, working memory refers to the cognitive system involved in maintaining and manipulating task-relevant information while a task is performed (Miyake & Shah, 1999). Working-memory capacity describes the notion that working memory is a limited capacity system: it has finite resources for processing and storage, and each person has a different resource pool that can be drawn from to successfully perform a task (Kane & Engle, 2002, 2003). To measure working memory capacity, one popular task (the operation span task) involves participants solving arithmetic problems while also trying to memorize words (Turner & Engle, 1989). In this manner, the task demands working-memory storage (to memorize the words) while also processing distracting arithmetic problems. The ability to maintain performance on a task in the face of distraction is a hallmark characteristic of individuals with high working-memory capacity. In our discussion of search errors, we noted that working memory may be critical for helping an observer maintain previously viewed features in memory while exploring the remainder of an image and associating subsequently identified features with features stored in working memory (Cain et al., 2013; Cain & Mitroff, 2013). In this case, higher working-memory capacity may be particularly important when there are multiple targets (rather than a single target) to be identified in an image. Furthermore, in our discussion of decision errors, we noted that some theories suggest that candidate hypotheses must be maintained in memory while evidence is accumulated during image inspection (Patel et al., 2005; Patel & Groen, 1986; Patel, Kaufman, & Arocha, 2002). Other theories suggest that hypotheses are formed early on and then tested during image inspection (Ledley & Lusted, 1959); it is important to point out that novices and experts may reason very differently during case interpretation, and one or both of these approaches may prove appropriate for

different observers. Some research demonstrates that individual differences in working memory capacity predict hypothesis generation and verification processes in a task involving customer order predictions (Dougherty & Hunter, 2003). Thus, in both search and decision-making there appear to be critical roles for working-memory capacity in predicting clinician performance. This possibility has not yet been examined in the context of medical image interpretation and diagnosis, and it is unclear how working-memory capacity might influence clinician eye movements, though it is an exciting direction for future research.

In our review of the literature, we also noted that most studies using eye tracking during medical image interpretation use static images. These include lung x-rays, histology slides, and skin lesions. This is not entirely surprising, as many medical images are indeed static, and interpreting eye movements over dynamic scenes can be very complex and time-consuming (Jacob & Karn, 2003; Jarodzka, Scheiter, Gerjets, & van Gog, 2010). There are also cases where images that are usually navigated (panned, zoomed) are artificially restricted, increasing the risk that results are no longer relevant to routine clinical practice. As modern technologies emerge in diagnostic medicine, this disconnect becomes increasingly disadvantageous. Indeed, many medical images are becoming more complex and dynamic; for example, interpreting live and replayed coronary angiograms, simulated dynamic patients during training, or navigating multiple layers of volumetric chest x-rays (Drew, Vö, & Wolfe, 2013; Rubin, 2015). Continued innovations in software for integrating dynamic visual scenes and eye movements will enable this type of research: for instance techniques that parse dynamic video stimuli based on navigation behavior (pause, rewind, play) to identify critical video frames (Yu, Ma, Nahrstedt, & Zhang, 2003). Some other techniques are being developed to provide rudimentary tagging and tracking of identifiable objects in a scene (Steciuk & Zwierno, 2015); such a technique might prove valuable for tracking a region of diagnostic interest that moves across a scene during playback (e.g., during coronary angiogram review).

It is also worth pointing out that many hospitals are introducing mandatory consultative expert second opinions for quality assurance purposes. For instance, Johns Hopkins Hospital and the University of Iowa Hospitals and Clinics introduced mandatory second opinions for surgical pathology (Kronz, Westra, & Epstein, 1999; Manion, Cohen, & Weydert, 2008). Not only are these mandates seen as valuable for the institutions involved (e.g., for reducing malpractice suits), but clinicians also perceive them as important for improving diagnostic accuracy (Geller et al., 2014). However, having an earlier physician's interpretation available during diagnosis may

unintentionally bias the second physician's diagnostic process. Indeed even a subtle probabilistic cue (e.g., a red dot that suggests an upcoming image contains a blast cell) can produce response bias in experienced diagnosticians (Trueblood et al., 2018). Thus, while viewing an expert's behavior may prove advantageous in certain conditions, future research must isolate the parameters that may dictate its success and balance the potential trade-off between guiding eye movements and potentially biasing interpretation. Furthermore, second opinions can also induce diagnostic disagreements among expert clinicians and necessitate time and expense for resolving disagreement and reaching a consensus diagnosis. Eye tracking may prove to be an invaluable arbiter for these sorts of disputes, allowing consultative physicians to view the eye movements of the physician who rendered the primary diagnosis. This practice may assist in helping the consultative physician understand which features were focused on, which features were missed, and understanding how the original physician arrived at their interpretation. Eye tracking could thus augment traditional text annotations to allow consultative physicians to see the case "through the eyes" of the other physician, possibly reducing disagreement or facilitating consensus through shared understanding. Similar strategies might be applied to peer cohorts or medical students and residents, allowing them to learn from each other's search patterns and successes and failures. On the other hand, this approach could introduce bias in the second physician and unintentionally increase agreement; if the first physician arrived at an incorrect interpretation, such agreement could be detrimental, demonstrating the importance of continuing research in this regard (Gandomkar, Tay, Brennan, Kozuch, & Mello-Thoms, 2018).

Conclusion

Medical image interpretation is a highly complex skill that influences not only diagnostic interpretations but also patient quality of life and survivability. Eye tracking is an innovative tool that is becoming increasingly commonplace in medical research and holds the potential to revolutionize trainee and clinician experiences.

Abbreviations

ADH: Atypical ductal hyperplasia; CBME: Competency-based medical education; DCIS: Ductal carcinoma in situ; EMME: Eye-movement modeling examples; LC-NE: Locus coeruleus-norepinephrine; ROI: Region of interest; SMI REDm: SensoMotoric Instruments' Remote Eye-tracking Device – mobile; VNPI: Van Nuys Prognostic Indicator

Acknowledgements

Not applicable.

Funding

This review was supported by funding from the National Cancer Institute of the National Institutes of Health under award numbers RO1 CA201376 and

RO1 CA225585. The content is solely the responsibility of the authors and does not necessarily represent the views of the National Cancer Institute or the National Institutes of Health.

Availability of data and materials

Not applicable.

Authors' contributions

TB conceived the review and prepared the manuscript, with critical revisions and feedback from authors JE, TD, and DW. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Center for Applied Brain and Cognitive Sciences, Tufts University, 200 Boston Ave., Suite 3000, Medford, MA 02155, USA. ²Department of Psychology, University of Utah, 380 1530 E, Salt Lake City, UT 84112, USA. ³Department of Pathology and University of Vermont Cancer Center, University of Vermont, 111 Colchester Ave., Burlington, VT 05401, USA. ⁴Department of Medicine, David Geffen School of Medicine at UCLA, University of California at Los Angeles, 10833 Le Conte Ave., Los Angeles, CA 90095, USA.

Received: 5 November 2018 Accepted: 1 February 2019

Published online: 22 February 2019

References

- Aggarwal, R., & Darzi, A. (2006). Technical-skills training in the 21st century. *New England Journal of Medicine*, 355, 2695–2696. <https://doi.org/10.1056/NEJMe068179>.
- Al-Moteri, M. O., Symmons, M., Plummer, V., & Cooper, S. (2017). Eye tracking to investigate cue processing in medical decision-making: a scoping review. *Computers in Human Behavior*, 66, 52–66. <https://doi.org/10.1016/j.chb.2016.09.022>.
- Ashraf, H., Sodergren, M. H., Merali, N., Mylonas, G., Singh, H., & Darzi, A. (2018). Eye-tracking technology in medical education: a systematic review. *Medical Teacher*, 40(1), 62–69. <https://doi.org/10.1080/0142159X.2017.1391373>.
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450. <https://doi.org/10.1146/annurev.neuro.28.061604.135709>.
- Ball, L. J., & Litchfield, D. (2017). Interactivity and embodied cues in problem solving, learning and insight: Further contributions to a "theory of hints". In *Cognition beyond the brain: computation, interactivity and human artifice*, (2nd ed., pp. 115–132). https://doi.org/10.1007/978-3-319-49115-8_6.
- Balslev, T., Jarodzka, H., Holmqvist, K., De Grave, W., Mujijs, A. M. M., Eika, B., ... Scherpbier, A. J. J. A. (2012). Visual expertise in paediatric neurology. *European Journal of Paediatric Neurology*, 16, 161–166. <https://doi.org/10.1016/j.ejpn.2011.07.004>.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637. <https://doi.org/10.1037/0033-2909.128.4.612>.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91, 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>.
- Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. *Cognition*, 106, 894–912. <https://doi.org/10.1016/j.cognition.2007.05.004>.

- Berbaum, K. S., Brandser, E. A., Franken, E. A., Dorfman, D. D., Caldwell, R. T., & Krupinski, E. A. (2001). Gaze dwell times on acute trauma injuries missed because of satisfaction of search. *Academic Radiology*, 8, 304–314. [https://doi.org/10.1016/S1076-6332\(03\)80499-3](https://doi.org/10.1016/S1076-6332(03)80499-3).
- Berbaum, K. S., Franken, E. A., Dorfman, D. D., Rooholamini, S. A., Kathol, M. H., Barloon, T. J., ... El-Khoury, G. Y. (1990). Satisfaction of search in diagnostic radiology. *Investigative Radiology*, 25, 133–140. <https://doi.org/10.1097/00004424-199002000-00006>.
- Berbaum, K. S., Krupinski, E. A., Schartz, K. M., Caldwell, R. T., Madsen, M. T., Hur, S., ... Franken, E. A. (2015). Satisfaction of search in chest radiography 2015. *Academic Radiology*, 22, 1457–1465. <https://doi.org/10.1016/j.acra.2015.07.011>.
- Bligh, J., Prideaux, D., & Parsell, G. (2001). PRISMS: new educational strategies for medical education. *Medical Education*, 35, 520–521. <https://doi.org/10.1046/j.1365-2923.2001.00984.x>.
- Blondon, K., Wipfli, R., & Lovis, C. (2015). Use of eye-tracking technology in clinical reasoning: a systematic review. *Studies in Health Technology and Informatics*, 210, 90–94. <https://doi.org/10.3233/978-1-61499-512-8-90>.
- Bok, H. G. J., Jaarsma, D. A. D. C., Spruijt, A., Van Beukelen, P., Van Der Vleuten, C. P. M., & Teunissen, P. W. (2016). Feedback-giving behaviour in performance evaluations during clinical clerkships. *Medical Teacher*, 38, 88–95. <https://doi.org/10.3109/0142159X.2015.1017448>.
- Bok, H. G. J., Teunissen, P. W., Spruijt, A., Fokkema, J. P. I., van Beukelen, P., Jaarsma, D. A. D. C., & van der Vleuten, C. P. M. (2013). Clarifying students' feedback-seeking behaviour in clinical clerkships. *Medical Education*, 47, 282–291. <https://doi.org/10.1111/medu.12054>.
- Bond, R. R., Zhu, T., Finlay, D. D., Drew, B., Kliffeld, P. D., Guldenring, D., ... Clifford, G. D. (2014). Assessing computerized eye tracking technology for gaining insight into expert interpretation of the 12-lead electrocardiogram: an objective quantitative approach. *Journal of Electrocardiology*, 47, 895–906. <https://doi.org/10.1016/j.jelectrocard.2014.07.011>.
- Bordage, G. (1999). Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Academic Medicine*, 74, S138. <https://doi.org/10.1097/00001888-199910000-00065>.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: brain, mind, experience, and school*. Committee on Learning Research and Educational Practice. Washington, D.C.: National Academy Press. [https://doi.org/10.1016/0885-2014\(91\)90049-J](https://doi.org/10.1016/0885-2014(91)90049-J).
- Brehmer, B. (1992). Dynamic decision making: human control of complex systems. *Acta Psychologica*, 81, 211–241. [https://doi.org/10.1016/0001-6918\(92\)90019-A](https://doi.org/10.1016/0001-6918(92)90019-A).
- Brunyé, T. T., Carney, P. A., Allison, K. H., Shapiro, L. G., Weaver, D. L., & Elmore, J. G. (2014). Eye movements as an index of pathologist visual expertise: a pilot study. *PLoS One*, 9(8). <https://doi.org/10.1371/journal.pone.0103447>.
- Brunyé, T. T., & Gardony, A. L. (2017). Eye tracking measures of uncertainty during perceptual decision making. *International Journal of Psychophysiology*, 120, 60–68. <https://doi.org/10.1016/j.ijpsycho.2017.07.008>.
- Brunyé, T. T., Haga, Z. D., Houck, L. A., & Taylor, H. A. (2017). You look lost: understanding uncertainty and representational flexibility in navigation. In J. M. Zacks, & H. A. Taylor (Eds.), *Representations in mind and world: essays inspired by Barbara Tversky*, (pp. 42–56). New York: Routledge. <https://doi.org/10.4324/9781315169781>.
- Brunyé, T. T., Mercan, E., Weaver, D. L., & Elmore, J. G. (2017). Accuracy is in the eyes of the pathologist: the visual interpretive process and diagnostic accuracy with digital whole slide images. *Journal of Biomedical Informatics*, 66, 171–179.
- Buettner, R. (2013). Cognitive workload of humans using artificial intelligence systems: towards objective measurement applying eye-tracking technology. *Lecture Notes in Computer Science*, 8077, 37–48. <https://doi.org/10.1007/978-3-642-40942-4-4>.
- Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2013). A taxonomy of errors in multiple-target visual search. *Visual Cognition*, 21, 899–921. <https://doi.org/10.1080/13506285.2013.843627>.
- Cain, M. S., & Mitroff, S. R. (2013). Memory for found targets interferes with subsequent performance in multiple-target visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 1398–1408. <https://doi.org/10.1037/a0030726>.
- Cain, M. S., Vul, E., Clark, K., & Mitroff, S. R. (2012). A Bayesian optimal foraging model of human visual search. *Psychological Science*, 23, 1047–1054. <https://doi.org/10.1177/0956797612440460>.
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cerebral Cortex*, 15, 1243–1249. <https://doi.org/10.1093/cercor/bhi007>.
- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, 16, 1905–1910. <https://doi.org/10.1016/j.cub.2006.07.065>.
- Carmody, D. P., Nodine, C. F., & Kundel, H. L. (1980). An analysis of perceptual and cognitive factors in radiographic interpretation. *Perception*, 9, 339–344. <https://doi.org/10.1068/p090339>.
- Chetwood, A. S. A., Kwok, K. W., Sun, L. W., Mylonas, G. P., Clark, J., Darzi, A., & Yang, G. Z. (2012). Collaborative eye tracking: a potential training tool in laparoscopic surgery. *Surgical Endoscopy and Other Interventional Techniques*, 26, 2003–2009. <https://doi.org/10.1007/s00464-011-2143-x>.
- Chun, M. M., & Wolfe, J. M. (1996). Just say no: how are visual searches terminated when there is no target present? *Cognitive Psychology*, 30, 39–78.
- Crowley, R. S., Naus, G. J., Stewart, J., & Friedman, C. P. (2003). Development of visual diagnostic expertise in pathology—an information-processing study. *Journal of the American Medical Informatics Association: JAMIA*, 10(1), 39–51. <https://dx.doi.org/10.1197%2Fjamia.M1123>. Accessed 1 Feb 2019.
- Custers, E. J. F. M. (2015). Thirty years of illness scripts: theoretical origins and practical applications. *Medical Teacher*, 37, 457–462. <https://doi.org/10.3109/0142159X.2014.956052>.
- Danziger, S., Kingstone, A., & Snyder, J. J. (1998). Inhibition of return to successively stimulated locations in a sequential visual search paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1467–1475. <https://doi.org/10.1037/0096-1523.24.5.1467>.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: guidelines for research and design. *Educational Psychology Review*, 21, 113–140. <https://doi.org/10.1007/s10648-009-9098-7>.
- Di Stasi, L. L., Catena, A., Cañas, J. J., Macknik, S. L., & Martinez-Conde, S. (2013). Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience and Biobehavioral Reviews*, 37(5), 968–975. <https://doi.org/10.1016/j.neubiorev.2013.03.011>.
- Dougherty, M. R. P., & Hunter, J. E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica*, 112, 263–282. [https://doi.org/10.1016/S0001-6918\(03\)00033-7](https://doi.org/10.1016/S0001-6918(03)00033-7).
- Drew, T., Vö, M. L. H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: sustained inattention blindness in expert observers. *Psychological Science*, 24, 1848–1853. <https://doi.org/10.1177/0956797613479386>.
- Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine*. New York: The Free Press.
- Eggin, T. K. P., & Feinstein, A. R. (1996). Context bias: a problem in diagnostic radiology. *JAMA*, 276, 1752–1755.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: an analysis of clinical reasoning*. Cambridge: Harvard University Press.
- Evans, K. K., Birdwell, R. L., & Wolfe, J. M. (2013). If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening. *PLoS One*, 8, 64366. <https://doi.org/10.1371/journal.pone.0064366>.
- Fabio, R. A., Incorpora, C., Errante, A., Mohammadhasni, N., Capri, T., Carrozza, C., ... Falzone, A. (2015). The influence of cognitive load and amount of stimuli on entropy through eye tracking measures. In *EuroAsianPacific Joint Conference on Cognitive Science*.
- Findlay, J. M., & Gilchrist, I. D. (2008). *Active vision: the psychology of looking and seeing*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198524793.001.0001>.
- Frank, J. R., & Danoff, D. (2007). The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Medical Teacher*, 29, 642–647. <https://doi.org/10.1080/01421590701746983>.
- Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., & Menke, L. (2016). Which eye tracker is right for your research? Performance evaluation of several cost variant eye trackers. In *Proceedings of the Human Factors and Ergonomics Society*, (pp. 1239–1243). <https://doi.org/10.1177/1541931213601289>.
- Gandomkar, Z., Tay, K., Brennan, P. C., Kozuch, E., & Mello-Thoms, C. R. (2018). Can eye-tracking metrics be used to better pair radiologists in a mammogram reading task? *Medical Physics*, 45, 4844–4956.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552. <https://doi.org/10.1007/s10648-011-9174-7>.
- Geller, B. M., Nelson, H. D., Carney, P. A., Weaver, D. L., Onega, T., Allison, K. H., ... Elmore, J. G. (2014). Second opinion in breast pathology: policy, practice and perception. *Journal of Clinical Pathology*, 67(11), 955–960. <https://doi.org/10.1136/jclinpath-2014-202290>.

- Gilhooly, K. J. (1990). Cognitive psychology and medical diagnosis. *Applied Cognitive Psychology*, 4, 261–272. <https://doi.org/10.1002/acp.2350040404>.
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, 10(2), 252–269. <https://doi.org/10.3758/CABN.10.2.252>.
- Giovinco, N. A., Sutton, S. M., Miller, J. D., Rankin, T. M., Gonzalez, G. W., Najafi, B., & Armstrong, D. G. (2015). A passing glance? Differences in eye tracking and gaze patterns between trainees and experts reading plain film bunion radiographs. *The Journal of Foot and Ankle Surgery: Official Publication of the American College of Foot and Ankle Surgeons*, 54(3), 382–391. <https://doi.org/10.1053/j.fjas.2014.08.013>.
- Green, M. L., Aagaard, E. M., Caverzagie, K. J., Chick, D. A., Holmboe, E., Kane, G., ... lobst, W. (2009). Charting the road to competence: developmental milestones for internal medicine residency training. *Journal of Graduate Medical Education*, 1, 5–20. <https://doi.org/10.4300/01.01.0003>.
- Gur, D., Sumkin, J. H., Rockette, H. E., Ganott, M., Hakim, C., Hardesty, L., ... Wallace, L. (2004). Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *Journal of the National Cancer Institute*, 96, 185–190. <https://doi.org/10.1093/jnci/djh067>.
- Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: a survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, 478–500. <https://doi.org/10.1109/TPAMI.2009.30>.
- Harden, R. M., Sowden, S., & Dunn, W. R. (1984). Educational strategies in curriculum development: the SPICES model. *Medical Education*, 18, 284–297. <https://doi.org/10.1111/j.1365-2923.1984.tb01024.x>.
- Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, 9(6), 467–479. <https://doi.org/10.1038/nrn2374>.
- Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: an overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception*, (pp. 269–293). Oxford: Elsevier.
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin and Review*, 16, 850–856. <https://doi.org/10.3758/PBR.16.5.850>.
- Henneman, P. L., Fisher, D. L., Henneman, E. A., Pham, T. A., Mei, Y. Y., Talati, R., ... Roche, J. (2008). Providers do not verify patient identity during computer order entry. *Academic Emergency Medicine*, 15, 641–648. <https://doi.org/10.1111/j.1553-2712.2008.00148.x>.
- Holmboe, E. S., Call, S., & Ficalora, R. D. (2016). Milestones and competency-based medical education in internal medicine. *JAMA Internal Medicine*, 176(11), 1601–1602.
- Holmboe, E. S., Edgar, L., & Hamstra, S. (2016). The milestones guidebook. Retrieved from <https://www.acgme.org/Portals/0/MilestonesGuidebook.pdf>. Accessed 1 Feb 2019.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: a comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: what it is and how to measure it. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, (pp. 45–52). <https://doi.org/10.1145/2168556.2168563>.
- Hong, S. K. (2005). Human stopping strategies in multiple-target search. *International Journal of Industrial Ergonomics*, 35, 1–12. <https://doi.org/10.1016/j.ergon.2004.06.004>.
- Ingre, M., Åkerstedt, T., Peters, B., Anund, A., & Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. *Journal of Sleep Research*, 15, 47–53. <https://doi.org/10.1111/j.1365-2869.2006.00504.x>.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506.
- Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research. Ready to deliver the promises. In *The mind's eye: cognitive and applied aspects of eye movement research*, (pp. 531–553). <https://doi.org/10.1016/B978-04451020-4/50031-1>.
- Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, 40, 815–827. <https://doi.org/10.1007/s11251-012-9218-5>.
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: how experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154. <https://doi.org/10.1016/j.learninstruc.2009.02.019>.
- Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70. <https://doi.org/10.1016/j.learninstruc.2012.11.004>.
- Juan, C.-H., Shorter-Jacobi, S. M., & Schall, J. D. (2004). Dissociation of spatial attention and saccade preparation. *Proceedings of the National Academy of Sciences*, 101, 15541–15544. <https://doi.org/10.1073/pnas.0403507101>.
- Jungk, A., Thull, B., Hoefl, A., & Rau, G. (2000). Evaluation of two new ecological interface approaches for the anesthesia workplace. *Journal of Ecological Monitoring and Computing*, 16, 243–258. <https://doi.org/10.1023/A:1011462726040>.
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin and Review*, 9, 637–671. <https://doi.org/10.3758/BF03196323>.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132(1), 47–70. <https://doi.org/10.1037/0096-3445.132.1.47>.
- Kassirer, J. P., Kopelman, R. I., & Wong, J. B. (1991). *Learning clinical reasoning*. Baltimore: Williams & Wilkins.
- Kern, D. E., Thomas, P. A., & Hughes, M. T. (1998). *Curriculum development for medical education: a six-step approach*. Baltimore: Johns Hopkins University Press.
- Khan, R. S. A., Tien, G., Atkins, M. S., Zheng, B., Panton, O. N. M., & Meneghetti, A. T. (2012). Analysis of eye gaze: do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? *Surgical Endoscopy and Other Interventional Techniques*, 26, 3536–3540. <https://doi.org/10.1007/s00464-012-2400-7>.
- Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science*, 8, 255–273. [https://doi.org/10.1016/S0364-0213\(84\)80003-8](https://doi.org/10.1016/S0364-0213(84)80003-8).
- Kirsh, D. (2009). Problem solving and situated cognition. In P. Robbins, & M. Aydede (Eds.), *The Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press.
- Kogan, J. R., Conforti, L., Bernabeo, E., lobst, W., & Holmboe, E. (2011). Opening the black box of clinical skills assessment via observation: a conceptual model. *Medical Education*, 45, 1048–1060. <https://doi.org/10.1111/j.1365-2923.2011.04025.x>.
- Kronz, J. D., Westra, W. H., & Epstein, J. I. (1999). Mandatory second opinion surgical pathology at a large referral hospital. *Cancer*, 86, 2426–2435. [https://doi.org/10.1002/\(SICI\)1097-0142\(19991201\)86:11<2426::AID-CNCR34>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0142(19991201)86:11<2426::AID-CNCR34>3.0.CO;2-3).
- Krupinski, E. A. (2005). Visual search of mammographic images: influence of lesion subtlety. *Academic Radiology*, 12(8), 965–969. <https://doi.org/10.1016/j.jacr.2005.03.071>.
- Krupinski, E. A., Berbaum, K. S., Scharzt, K. M., Caldwell, R. T., & Madsen, M. T. (2017). The impact of fatigue on satisfaction of search in chest radiography. *Academic Radiology*, 24, 1058–1063. <https://doi.org/10.1016/j.jacr.2017.03.021>.
- Krupinski, E. A., Graham, A. R., & Weinstein, R. S. (2013). Characterizing the development of visual search experience in pathology residents viewing whole slide images. *Human Pathology*, 44, 357–364.
- Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J., Bhattacharyya, A. K., Scott, K. M., ... Weinstein, R. S. (2006). Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. *Human Pathology*, 37(12), 1543–1556.
- Kundel, H. L., & La Follette, P. S. (1972). Visual search patterns and experience with radiological images. *Radiology*, 103, 523–528.
- Kundel, H. L., & Nodine, C. F. (1978). Studies of eye movements and visual search in radiology. In J. W. Senders, D. F. Fisher, & R. A. Monty (Eds.), *Eye movements and the higher psychological processes*, (pp. 317–327). Hillsdale: Lawrence Erlbaum Associates.
- Kundel, H. L., & Nodine, C. F. (2010). A short history of image perception in medical radiology. In E. Samei, & E. A. Krupinski (Eds.), *The handbook of medical image perception and techniques*, (pp. 9–20). London: Cambridge University Press.
- Kundel, H. L., Nodine, C. F., & Krupinski, E. A. (1990). Computer-displayed eye position as a visual aid to pulmonary nodule interpretation. *Investigative Radiology*, 25, 890–896. <https://doi.org/10.1097/00004424-199008000-00004>.
- Kundel, H. L., Nodine, C. F., Krupinski, E. A., & Mello-Thoms, C. (2008). Using gaze-tracking data and mixture distribution analysis to support a holistic model

- for the detection of cancers on mammograms. *Academic Radiology*, 15(7), 881–886. <https://doi.org/10.1016/j.acra.2008.01.023>.
- Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: a window to the preconscious? *Perspectives on Psychological Science*, 7(1), 18–27. <https://doi.org/10.1177/1745691611427305>.
- Ledley, R. S., & Lusted, L. B. (1959). Reasoning foundations of medical diagnosis. *Science*, 130, 9–21. <https://doi.org/10.1126/science.130.3366.9>.
- Leff, D. R., James, D. R. C., Orihuela-Espina, F., Kwok, K.-W., Sun, L. W., Mylonas, G., ... Yang, G.-Z. (2015). The impact of expert visual guidance on trainee visual search strategy, visual attention and motor skills. *Frontiers in Human Neuroscience*, 9, 526. <https://doi.org/10.3389/fnhum.2015.00526>.
- Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing x-ray pictures. In *The nature of expertise*, (pp. 311–342). Hillsdale: Lawrence Erlbaum Associates.
- Lester, S. C., & Hicks, D. (2016). *Diagnostic pathology: breast*, (2nd ed.,). Philadelphia: Elsevier.
- Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied*, 16, 251–262. <https://doi.org/10.1037/a0020082>.
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6–14.
- Lobmaier, J. S., Fischer, M. H., & Schwaninger, A. (2006). Objects capture perceived gaze direction. *Experimental Psychology*, 53, 117–122. <https://doi.org/10.1027/1618-3169.53.2.117>.
- Lundgrén-Laine, H., & Salanterä, S. (2010). Think-aloud technique and protocol analysis in clinical decision-making research. *Qualitative Health Research*, 20, 565–575. <https://doi.org/10.1177/1049732309354278>.
- Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology*, 121, 62–73. <https://doi.org/10.1016/j.biopsycho.2016.10.002>.
- Manion, E., Cohen, M. B., & Weydert, J. (2008). Mandatory second opinion in surgical pathology referral material: clinical consequences of major disagreements. *American Journal of Surgical Pathology*, 32, 732–737. <https://doi.org/10.1097/PAS.0b013e31815a04f5>.
- Manning, D., Ethell, S., Donovan, T., & Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography*, 12(2), 134–142. <https://doi.org/10.1016/j.radi.2005.02.003>.
- Martin, C., Cegarra, J., & Averty, P. (2011). Analysis of mental workload during en-route air traffic control task execution based on eye-tracking technique. In *Lecture notes in computer science (including subseries Lecture notes in artificial intelligence and Lecture notes in bioinformatics)*, (pp. 592–597). https://doi.org/10.1007/978-3-642-21741-8_63.
- Martinez-Conde, S., Otero-Millan, J., & MacKnik, S. L. (2013). The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, 14, 83–96. <https://doi.org/10.1038/nrn3405>.
- Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of integrative reading of an illustrated text: effects on processing and learning. *Contemporary Educational Psychology*, 41, 172–187. <https://doi.org/10.1016/j.cedpsych.2015.01.004>.
- Matsumoto, H., Terao, Y., Yugeta, A., Fukuda, H., Emoto, M., Furubayashi, T., ... Ugawa, Y. (2011). Where do neurologists look when viewing brain CT images? An eye-tracking study involving stroke cases. *PLoS One*, 6, e28928. <https://doi.org/10.1371/journal.pone.0028928>.
- McCarley, J. S., & Carruth, D. (2004). Oculomotor scanning and target recognition in luggage x-ray screening. *Cognitive Technology*, 9, 26–29.
- McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual skills in airport-security screening. *Psychological Science*, 15(5), 302–306. <https://doi.org/10.1111/j.0956-7976.2004.00673.x>.
- Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., ... Maiz, G. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Academic Radiology*, 12(7), 830–840. <https://doi.org/10.1016/j.acra.2005.03.068>.
- Mercan, E., Aksoy, S., Shapiro, L. G., Weaver, D. L., Brunyé, T. T., & Elmore, J. G. (2016). Localization of diagnostically relevant regions of interest in whole slide images: a comparative study. *Journal of Digital Imaging*, 29, 496–506. <https://doi.org/10.1007/s10278-016-9873-1>.
- Mercan, E., Shapiro, L. G., Brunyé, T. T., Weaver, D. L., & Elmore, J. G. (2017). Characterizing diagnostic search patterns in digital breast pathology: scanners and drillers. *Journal of Digital Imaging*. <https://doi.org/10.1007/s10278-017-9990-5>.
- Meyberg, S., Werkle-Bergner, M., Sommer, W., & Dimigen, O. (2015). Microsaccade-related brain potentials signal the focus of visuospatial attention. *NeuroImage*, 104, 79–88. <https://doi.org/10.1016/j.neuroimage.2014.09.065>.
- Miyake, A., & Shah, P. (1999). Models of working memory: an introduction. In A. Miyake, & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control*, (pp. 1–27). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139174909>.
- Montagnini, A., & Chelazzi, L. (2005). The urgency to look: prompt saccades to the benefit of perception. *Vision Research*, 45, 3391–3401. <https://doi.org/10.1016/j.visres.2005.07.013>.
- Motowildo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, 10, 71–83. https://doi.org/10.1207/s15327043hup1002_1.
- Nabil, N. M., Guemei, A. A., Alsaaid, H. F., Salem, R., Abuzaid, L. Z., Abdulhalim, D., ... Al-Kaaba, A. F. (2013). Impact of introducing a diagnostic scheme to medical students in a problem based learning curriculum. *Medical Science Educator*, 23(1), 16–26.
- Nagarkar, D. B., Mercan, E., Weaver, D. L., Brunyé, T. T., Carney, P. A., Rendi, M. H., ... Elmore, J. G. (2016). Region of interest identification and diagnostic agreement in breast pathology. *Modern Pathology*, 29(9), 1004. <https://doi.org/10.1038/modpathol.2016.85>.
- Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics*, 36, 289–300. <https://doi.org/10.1016/j.ergon.2005.11.008>.
- Nasca, T. J., Philibert, I., Brigham, T., & Flynn, T. C. (2012). The next GME accreditation system—rationale and benefits. *New England Journal of Medicine*, 366, 1051–1056. <https://doi.org/10.1056/NEJMs1200117>.
- Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin and Review*, 17, 718–724. <https://doi.org/10.3758/PBR.17.5.718>.
- Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2015). *Straight choices: the psychology of decision making*. 2nd ed. London: Psychology Press. <https://doi.org/10.4324/9781315727080>.
- Nodine, C., & Kundel, H. (1987). Using eye movements to study visual search and to improve tumor detection. *Radiographics*, 7, 1241–1250. <https://doi.org/10.1148/radiographics.7.6.3423330>.
- O'Meara, P., Munro, G., Williams, B., Cooper, S., Bogossian, F., Ross, L., ... McClounan, M. (2015). Developing situation awareness amongst nursing and paramedicine students utilizing eye tracking technology and video debriefing techniques: a proof of concept paper. *International Emergency Nursing*, 23, 94–99. <https://doi.org/10.1016/j.ijen.2014.11.001>.
- O'Neill, E. C., Kong, Y. X. G., Connell, P. P., Ong, D. N., Haymes, S. A., Coote, M. A., & Crowston, J. G. (2011). Gaze behavior among experts and trainees during optic disc examination: does how we look affect what we see? *Investigative Ophthalmology and Visual Science*, 52, 3976–3983. <https://doi.org/10.1167/iov.10-6912>.
- Patel, V. L., Arocha, J., & Zhang, J. (2005). Thinking and reasoning in medicine. In *The Cambridge handbook of thinking and reasoning*, (pp. 727–750). Cambridge: Cambridge University Press.
- Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10, 91–116. [https://doi.org/10.1016/S0364-0213\(86\)80010-6](https://doi.org/10.1016/S0364-0213(86)80010-6).
- Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2002). Emerging paradigms of cognition in medical decision-making. *Journal of Biomedical Informatics*, 35, 52–75.
- Pinnock, R., Young, L., Spence, F., & Henning, M. (2015). Can think aloud be used to teach and assess clinical reasoning in graduate medical education? *Journal of Graduate Medical Education*, 7, 334–337. <https://doi.org/10.4300/JGME-D-14-00601.1>.
- Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision*, 10(10), 3. <https://doi.org/10.1167/10.10.3>.
- Ratwani, R. M., & Trafton, J. G. (2011). A real-time eye tracking system for predicting and preventing postcompletion errors. *Human-Computer Interaction*, 26, 205–245. <https://doi.org/10.1080/07370024.2011.601692>.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>.

- Rich, A. N., Kunar, M. A., Van Wert, M. J., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision*, 8, 15. <https://doi.org/10.1167/8.15.15>.
- Richstone, L., Schwartz, M. J., Seideman, C., Cadeddu, J., Marshall, S., & Kavoussi, L. R. (2010). Eye metrics as an objective assessment of surgical skill. *Annals of Surgery*, 252(1), 177–182. <https://doi.org/10.1097/SLA.0b013e3181e464fb>.
- Roads, B. D., Xu, B., Robinson, J. K., & Tanaka, J. W. (2018). The easy-to-hard training advantage with real-world medical images. *Cognitive Research: Principles and Implications*, 3, 38.
- Rubin, G. D. (2015). Lung nodule and cancer detection in CT screening. *Journal of Thoracic Imaging*, 30, 130–138.
- Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). Use of eye movements as feedforward training for a synthetic aircraft inspection task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI '05*, (pp. 289–300). <https://doi.org/10.1145/1054972.1054993>.
- Samuel, S., Kundel, H. L., Nodine, C. F., & Toto, L. C. (1995). Mechanism of satisfaction of search: eye position recordings in the reading of chest radiographs. *Radiology*, 194, 895–902.
- Sibbald, M., de Bruin, A. B. H., Yu, E., & van Merriënboer, J. J. G. (2015). Why verifying diagnostic decisions with a checklist can help: insights from eye tracking. *Advances in Health Sciences Education*, 20, 1053–1060. <https://doi.org/10.1007/s10459-015-9585-1>.
- Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45(5), 679–687. <https://doi.org/10.1111/j.1469-8986.2008.00681.x>.
- Simon, H. A. (1983). Why should machines learn? In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning*, (pp. 25–37). Los Altos: Morgan Kaufmann. https://doi.org/10.1007/978-3-662-12405-5_2.
- Simpson, J. G., Furnace, J., Crosby, J., Cumming, A. D., Evans, P. A., Ben David, F., ... Macpherson, S. G. (2002). The Scottish doctor—Learning outcomes for the medical undergraduate in Scotland: a foundation for competent and reflective practitioners. *Medical Teacher*, 24, 136–143. <https://doi.org/10.1080/01421590220120713>.
- Smith, M. J. (1967). *Error and variation in diagnostic radiology*. Springfield: C. C. Thomas.
- Sox, H. C., Blatt, M. A., Higgins, M. C., & Marton, K. I. (1988). *Medical decision making*. Boston: Butterworths. <https://doi.org/10.1002/9781118341544>.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning Memory and Cognition*, 24, 1521–1543. <https://doi.org/10.1037/0278-7393.24.6.1521>.
- Steciuk, H., & Zwierno, T. (2015). Gaze behavior in basketball shooting: preliminary investigations. *Trends in Sport Sciences*, 22, 89–94.
- Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. In *Proceedings of the 6th International Conference on Multimodal Interfaces—ICMI '04*, (pp. 9–15). <https://doi.org/10.1145/1027933.1027936>.
- Sumner, P. (2011). Determinants of saccade latency. In *The Oxford handbook of eye movements*, (pp. 413–424). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199539789.013.0022>.
- Sunday, M. A., Donnelly, E., & Gauthier, I. (2017). Individual differences in perceptual abilities in medical imaging: the Vanderbilt Chest Radiograph Test. *Cognitive Research: Principles and Implications*, 2, 36. <https://doi.org/10.1186/s41235-017-0073-4>.
- Swing, S. R. (2007). The ACGME outcome project: retrospective and prospective. *Medical Teacher*, 29, 648–654. <https://doi.org/10.1080/01421590701392903>.
- Swing, S. R., Beeson, M. S., Carraccio, C., Coburn, M., Iobst, W., Selden, N. R., ... Vydareny, K. (2013). Educational milestone development in the first 7 specialties to enter the next accreditation system. *Journal of Graduate Medical Education*, 5, 98–106. <https://doi.org/10.4300/JGME-05-01-33>.
- Szulewski, A., Braund, H., Egan, R., Hall, A. K., Dagnone, J. D., Gegenfurtner, A., & van Merriënboer, J. J. G. (2018). Through the learner's eyes: eye-tracking augmented debriefing in medical simulation. *Journal of Graduate Medical Education*, 10(3), 340–341.
- Thomas, L. E., & Lleras, A. (2009a). Covert shifts of attention function as an implicit aid to insight. *Cognition*, 111, 168–174. <https://doi.org/10.1016/j.cognition.2009.01.005>.
- Thomas, L. E., & Lleras, A. (2009b). Swinging into thought: directed movement guides insight in problem solving. *Psychonomic Bulletin and Review*, 16, 719–723. <https://doi.org/10.3758/PBR.16.4.719>.
- Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G. Z., & Darzi, A. (2014). Eye tracking for skills assessment and training: a systematic review. *Journal of Surgical Research*, 191(1), 169–178. <https://doi.org/10.1016/j.jss.2014.04.032>.
- Tourassi, G. D., Mazurowski, M. A., Harrawood, B. P., & Krupinski, E. A. (2010). Exploring the potential of context-sensitive CAdE in screening mammography. *Medical Physics*, 37, 5728–5736. <https://doi.org/10.1118/1.3501882>.
- Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.
- Trueblood, J. S., Holmes, W. R., Seegmiller, A. C., Douds, J., Compton, M., Szentirmai, E., ... Eichbaum, Q. (2018). The impact of speed and bias on the cognitive processes of experts and novices in medical image decision-making. *Cognitive Research: Principles and Implications*, 3, 28.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154. [https://doi.org/10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5).
- Underwood, G., & Radach, R. (1998). Eye guidance and visual information processing. In *Eye guidance in reading and scene perception*, (pp. 1–27). New York: Elsevier. <https://doi.org/10.1016/B978-008043361-5/50002-X>.
- Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science*, 283(5401), 549–554. <https://doi.org/10.1126/science.283.5401.549>.
- van der Gijp, A., Ravesloot, C. J., Jarodzka, H., van der Schaaf, M. F., van der Schaaf, I. C., van Schaijk, J. P. J., & ten Cate, T. J. (2017). How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. *Advances in Health Sciences Education*, 22, 765–787. <https://doi.org/10.1007/s10459-016-9698-1>.
- van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25, 785–791. <https://doi.org/10.1016/j.chb.2009.02.007>.
- Velichkovsky, B. M. (1995). Communicating attention: gaze position transfer in cooperative problem solving. *Pragmatics & Cognition*, 3, 199–223. <https://doi.org/10.1075/pc.3.2.02vel>.
- Võ, M. L. H., Aizenman, A. M., & Wolfe, J. M. (2016). You think you know where you looked? You better look again. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 1477–1481. <https://doi.org/10.1037/xhp0000264>.
- Voisin, S., Pinto, F., Morin-Ducote, G., Hudson, K. B., & Tourassi, G. D. (2013). Predicting diagnostic error in radiology via eye-tracking and image analytics: preliminary investigation in mammography. *Medical Physics*, 40, 101906. <https://doi.org/10.1118/1.4820536>.
- Wald, H. S., Davis, S. W., Reis, S. P., Monroe, A. D., & Borkan, J. M. (2009). Reflecting on reflections: enhancement of medical education curriculum with structured field notes and guided feedback. *Academic Medicine*, 84(7), 830–837. <https://doi.org/10.1097/ACM.0b013e318a8592f>.
- Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: rare items often missed in visual searches. *Nature*, 435, 439. <https://doi.org/10.1038/435439a>.
- Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, 20, 121–124. <https://doi.org/10.1016/j.cub.2009.11.066>.
- Wood, G., Batt, J., Appelboam, A., Harris, A., & Wilson, M. R. (2014). Exploring the impact of expertise, clinical history, and visual search on electrocardiogram interpretation. *Medical Decision Making*, 34, 75–83. <https://doi.org/10.1177/0272989X13492016>.
- Xu-Wilson, M., Zee, D. S., & Shadmehr, R. (2009). The intrinsic value of visual information affects saccade velocities. *Experimental Brain Research*, 196, 475–481. <https://doi.org/10.1007/s00221-009-1879-1>.
- Young, L. R., & Stark, L. (1963). Variable feedback experiments testing a sampled data model for eye tracking movements. *IEEE Transactions on Human Factors in Electronics*, 1, 38–51. <https://doi.org/10.1109/THFE.1963.231285>.
- Yu, B., Ma, W.-Y., Nahrstedt, K., & Zhang, H.-J. (2003). Video summarization based on user log enhanced link analysis. In *Proceedings of the ACM Multimedia Conference (ACMMM)*, (pp. 382–391). Berkeley. <https://doi.org/10.1145/957013.957095>.
- Yuval-Greenberg, S., Merriam, E. P., & Heeger, D. J. (2014). Spontaneous microsaccades reflect shifts in covert attention. *Journal of Neuroscience*, 34, 13693–13700. <https://doi.org/10.1523/JNEUROSCI.0582-14.2014>.