American Psychological Association

Journal of Experimental Psychology: Applied

© 2020 American Psychological Association
ISSN: 1076-898X

2020, Vol. 26, No. 4, 579–592
http://dx.doi.org/10.1037/xap0000274

Seeing Isn't Necessarily Believing: Misleading Contextual Information Influences Perceptual-Cognitive Bias in Radiologists

Bradley Fawver, Joseph L. Thomas, Trafton Drew, Megan K. Mills, William F. Auffermann, Keith R. Lohse, and A. Mark Williams
University of Utah

A substantial number of medical errors in radiology are attributed to failures of perception or decision making, although it is believed that experience (or expertise) might buffer diagnosticians from some types of perceptual-cognitive bias. We examined how the quality of contextual information influences decision making and how underlying perceptual-cognitive processes change as a function of experience and diagnostic accuracy. Twenty-one radiologists dictated their findings on 16 deidentified musculoskeletal radiographic cases while wearing a mobile-eye tracking system. Patient histories were mismatched on a subset of cases to be miscued relative to the correct diagnosis. Experienced radiologists outperformed less-experienced participants, but no systematic differences in gaze behaviors emerged between groups. Miscued case notes increased perceptual-cognitive bias in both groups, resulting in an approximate 40% decrease in diagnostic accuracy. Most errors were judgment errors, meaning participants visually fixated on the abnormality for longer than a second yet still failed to make the correct diagnosis. Findings suggest a physician's confidence in their diagnosis might be misplaced after spending insufficient time extracting relevant information from key areas of the visual display, or when decisions are based primarily on a priori expectations derived from patient histories.

Public Significance Statement
This study demonstrates that the accuracy of case notes/patient histories play an important role in cueing radiologists. In addition, these data highlight the importance of spending sufficient time visually fixating on potential abnormalities to avoid perceptual-cognitive bias during diagnostic decision-making.

Keywords: case history, decision-making, diagnostic error, eye-tracking, verbal reports

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Recent reports indicate that medical errors pose a continuing risk to health across the world, as well as presenting a significant financial challenge (Makary & Daniel, 2016; Shojania & Dixon-Woods, 2017). Error rates in radiology were first recognized 70 years ago (Garland, 1949), yet their prevalence has not substantially decreased in the intervening period. The "real-time" error rate among radiologists has remained approximately 4% (Brady, Laoide, Mccarthy, & Mcdermott, 2012) and can be as high as 30–50% in some subspecialties, such as breast cancer imaging. Moreover, billions of dollars are awarded in malpractice lawsuits each year, with 75% of these claims attributed to diagnostic error (Berlin, 2007; Saber Tehrani et al., 2013). The significant variance in accuracy rates that exists across radiographic professionals remains largely unexplained (Beam, Conant, Sickles, & Weinstein, 2003), but it is believed that contextual factors might shape how radiologists perceive and process information from medical images. In the present study, we address a notable gap in the literature on diagnostic error by examining how case context, derived from the presence of accurately cued or miscued case notes, influences perceptual-cognitive bias and decision making in clinical radiographers of varying experience.

Graber, Franklin, and Gordon (2005) defined diagnostic errors as judgments that are incorrect or delayed until confirmed with a follow up test. The errors that occur in clinical radiology can be broadly grouped into failures in perception (i.e., detection errors), failures in decision making (i.e., judgment errors), failures in communication, or failures related to follow-up procedures (Krupinski, 2010). The overwhelming majority of diagnostic errors

This article was published Online First April 23, 2020.

Bradley Fawver and Joseph L. Thomas, Department of Health, Kinesiology, and Recreation, University of Utah; Trafton Drew, Department of Psychology, University of Utah; Megan K. Mills and William F. Auffermann, Department of Radiology and Imaging Sciences, University of Utah; Keith R. Lohse and A. Mark Williams, Department of Health, Kinesiology, and Recreation, University of Utah.

This study was funded by an RSNA/AUR/APDR/SCARD Radiology Education Research Development Grant (ERD1805). We would like to acknowledge Cullen Woodley for his assistance with data management and processing, verbal report coding, and manuscript editing. The data that support the findings of this study will be openly available at the time of acceptance in a repository at The Open Science Framework (link to data file: https://osf.io/zdp6r/).

Correspondence concerning this article should be addressed to Bradley Fawver, who is now at the Department of Physical Medicine and Rehabilitation, University of Utah, 383 Colorow Drive, Suite 260, Salt Lake City, UT 84108. E-mail: bfawver@health.utah.edu

(~70%) are believed to be perceptual in nature, meaning that the clinician fails to visually fixate on, or to pick up relevant information from, the area containing an abnormality (Rosenkrantz & Bansal, 2016). Conversely, decision-making errors are defined as occurring even after an abnormality or lesion is fixated on for a period of time (generally >1 s), yet still not identified or correctly evaluated (see Bruno, Walker, & Abujudeh, 2015).

The majority of researchers in the field of radiology have focused on identifying the causes of perceptual and decision making errors (e.g., Al-Moteri, Symmons, Plummer, & Cooper, 2017; Krupinski, 2011), with fewer efforts directed toward the multitiered communication process associated with a medical diagnosis (for an exception, see Siewert, Brook, Hochman, & Eisenberg, 2016). As a result, improvements in radiography methods, imaging techniques, and systems for identifying abnormalities (e.g., artificial networks, machine learning) do not always yield better patient outcomes (Bradley et al., 2015; Pfeffer et al., 2004), highlighting the prevailing problem associated with human judgment (Wolfe, Evans, Drew, Aizenman, & Josephs, 2016). Cognitive biases represent a deviation in judgment or rationality, which during visually based tasks often involves perceptual distortions (see Blumenthal-Barby & Krieger, 2015; Marewski & Gigerenzer, 2012); globally termed "perceptual-cognitive bias." A physician's intuitive process is intrinsically tied to their ability to make accurate diagnostic decisions (see Hall, 2002), making the development of perceptual-cognitive expertise (i.e., knowing where to look, why to look there, interpreting information correctly) a continuing priority in clinical radiography.

Over the last 20 years, numerous researchers have attempted to identify the processes and mechanisms underlying the ability of radiographers to effectively and efficiently make accurate diagnostic judgments. It appears that expert (or more-experienced) performers exhibit different visual search behaviors compared to nonexperts (e.g., Cooper, Gale, Darker, Toms, & Saada, 2009; Donovan & Manning, 2006), which likely contributes to their superior decision-making performance (for a meta-analysis, see Gegenfurtner, Lehtinen, & Säljö, 2011). For example, experts have demonstrated improved efficiency of visual search (e.g., Giovinco et al., 2015) and greater sensitivity to critical visual targets (e.g., Evans, Cohen, et al., 2011) compared to less-expert diagnosticians. Moreover, experts typically exhibit longer saccades (i.e., greater amplitude) during diagnostic tasks (Kocak, Ober, Berme, & Melvin, 2005; Krupinski et al., 2006), potentially indicating how their a priori expectation of what to look for drives the search process. Relatedly, experts demonstrate earlier fixations on key areas of interest (i.e., site of an abnormality) and fewer fixations on irrelevant areas (e.g., Wood et al., 2013), exhibit shorter fixation periods prior to a diagnostic judgment (e.g., Leong, Nicolaou, Emery, Darzi, & Yang, 2007), require less visual coverage of a medical image (e.g., Krupinski, 1996; Manning, Ethell, Donovan, & Crawford, 2006), and require fewer fixations on relevant areas of interest to make an accurate diagnosis (e.g., Bertram, Helle, Kaakinen, & Svedström, 2013; Krupinski et al., 2006; Manning et al., 2006). Overall, published reports indicate that expert diagnosticians show greater accuracy and consistency during visual search tasks (e.g., Leong et al., 2007; Wood et al., 2013), greater proficiency at interpreting medical images (e.g., Grünheid, Hollevoet, Miller, & Larson, 2013; Nodine, Mello-Thoms, Kundel, & Weinstein, 2002), require less time to make correct diagnoses (e.g., Cooper et al., 2010; Giovinco et al., 2015; Mallett, 2014), and are more confident in their judgments (e.g., Wood et al., 2013).

As per other domains, experts in a chosen subspecialty develop their skills by accumulating domain-specific knowledge through extensive practice over many years (Ericsson, 2015). Typically, radiologists complete at least a bachelor's degree, followed by 3 to 4 years of medical school, several years of residency, and a few additional years of specialized fellowship training. The prevailing assumption in many medical domains is that more-experienced individuals always perform tasks more efficiently than less-experienced individuals. The prototypical approach has been to model the processes of experts in order to translate changes in perception to changes in performance. However, recent findings suggests medical trainees develop "more expert" visual search strategies faster than they develop expert diagnostic decision-making skills (Kelly, Rainford, Darcy, Kavanagh, & Toomey, 2016). Moreover, diagnostic errors due to deficits in attention and perception or decision-making can often be traced to errors in the underlying cognitive processes, rather than gaze behaviors per se (Lee, Nagy, Weaver, & Newman-Toker, 2013), implying that improvements in systematic search methods may not necessarily improve judgment accuracy.

Modeling expertise has proven useful in numerous other domains (see Ericsson, Hoffman, Kozbelt, & Williams, 2018), but quantifying expert behavior and translating this knowledge into enhanced training methods has proven difficult (Gegenfurtner et al., 2017). One barrier in clinical medicine is that experience (i.e., time spent in a domain) does not always correlate with expertise (i.e., objective superior performance on a given task). For instance, experienced medical professionals often fail to make improved diagnoses or implement interventions that lead to enhanced treatment outcomes when compared with less-qualified and less-experienced professionals (e.g., Causer, Barach, & Williams, 2014; Ericsson & Lehmann, 1996). In some cases, the length of a clinician's professional experience can be unrelated, or even negatively related, to the quality of performance (see Choudhry, Fletcher, & Soumerai, 2005). The discrepancy between experience and expertise remains largely unexplained in diagnostic medicine, as well as the influence of experience on an individual's susceptibility to perceptual-cognitive bias. It may be the case that novices and trainees are more susceptible to attentional and perceptual inefficiencies due to their lack of knowledge in the domain, whereas more-experienced diagnosticians suffer from the opposite problem, namely, an overreliance on previous knowledge and contextual information (Kok, De Bruin, Robben, & Van Merriënboer, 2012; Summerfield & Egner, 2009).

The influence of context on decision making in medicine has been recognized for over 20 years (Egglin & Feinstein, 1996), yet only recently have researchers started to empirically study how factors such as imaging modality (Cooper et al., 2010) or prevalence bias (Evans, Tambouret, Evered, Wilbur, & Wolfe, 2011; Wolfe et al., 2007) influence the perceptual-cognitive processes underlying diagnostic judgments. Radiologists and pathologists are unique compared to other physicians in that they typically have no direct interaction with the patient whose images they are viewing. As with other professional domains, radiologists rely on heuristics (i.e., rules used to form judgments) to problem solve and make decisions. Heuristics are not

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

necessarily good or bad in terms of diagnostic outcomes (Wegwarth, Gaissmaier, & Gigerenzer, 2009), but an overreliance on rules can lead to cognitive bias and faulty decision-making processes (see Blumenthal-Barby & Krieger, 2015; Marewski & Gigerenzer, 2012). An unintended consequence of diagnostic experience is that individuals may develop an overreliance on case history and epidemiological information when making diagnostic judgments. As a result, a critical and often overlooked problem exists in this field; seasoned radiologists tend to make fewer errors on relatively easy cases compared to those who are less-experienced, yet their increased declarative knowledge and experience might make these clinicians increasingly susceptible to perceptual-cognitive bias during difficult or atypical cases where the contextual information provided about the case is incongruent with the underlying problem.

Any available contextual information about the case (i.e., likelihood of particular findings being present) is provided primarily through the consulting physician/provider case notes in the medical record and/or a brief indication for why the study was ordered. Therefore, one prominent source of context in diagnostic medicine comes from the patient's clinical history. Physicians should provide adequate clinical information so that technicians and radiologists can perform their jobs in a more focused manner (George, Espinosa, & Quattrone, 1992); however, in daily practice, case indications can often be vague, incomplete, or misleading. This problem is particularly relevant for diagnosticians in emergency settings where previous medical images may not be readily available and decisions must be made quickly with limited clinical information. Poor quality case notes can subsequently create a breeding ground for perceptual-cognitive and decision-making bias. For example, if case notes indicate a patient was just involved in a motor vehicle accident, the attending radiologist is likely to be more sensitive toward acute problems, potentially ignoring signs of chronic disease. Conversely, radiologists searching for signs of cancer may display "inattentional blindness" to unrelated, yet clinically relevant, findings such as a bone fracture (Drew, Võ, & Wolfe, 2013; Simons & Chabris, 1999).

Although the influence of context-specific information has been explored in other domains (e.g., Runswick, Roca, Williams, Bezodis, & North, 2018), few researchers have examined how context shapes decision-making performance in diagnostics (for a notable exception, see McRobert et al., 2013). Moreover, while published reports indicate that precise and accurate clinical case notes enhance diagnostic performance (Loy, 2004), to our knowledge, no work exists examining how the quality of case notes influences diagnostic accuracy in radiology using multiple process tracing measures of performance and groups of varying diagnostic experience.

We address these notable gaps in the literature by examining how the quality of contextual information, provided in the form of case notes/study indication, influences perceptual-cognitive processes during diagnostic imaging. Specifically, we asked more- and less-experienced radiologists to report their clinical findings on a set of musculoskeletal (MSK) cases that were presented with case notes of either high (i.e., correctly cueing the primary finding) or low quality (i.e., miscued to the actual diagnosis). Visual search behaviors were assessed using a head-mounted eye tracking system. While gaze recording is commonplace in studies of diagnostic behavior (see van der Gijp et al., 2017), to our knowledge only a few researchers have attempted to measure the underlying cognitive processes associated with diagnostic performance using self-report measures such as verbal reports (see Cormier, Pickett-Hauber, & Whyte, 2010; McRobert et al., 2013; McRobert, Mercer, Raw, Goulding, & Williams, 2017; Whyte, Cormier, & Pickett-Hauber, 2010). For example, McRobert and colleagues collected concurrent verbal reports from a group of skilled and less-skilled emergency medicine doctors who diagnosed emergency room scenarios. They coded verbal reports based on an accepted procedure (see McRobert, Ward, Eccles, & Williams, 2011) into cognitive, evaluative, predictive, and deep planning statements. The experts engaged in more predictive and deep planning thought processes compared to novices who lacked the requisite experience to make a priori judgments. In the present study, we extend this work to the field of radiology by collecting retrospective verbal reports after a subset of cases to examine the underlying thought processes and cognitive strategies used. Also, in an effort to provide some evidence (albeit exploratory) as to the link between existing process-tracing mechanisms and diagnostic performance, we included measures of self-reported confidence and mental effort.

We hypothesized that miscued case notes would negatively bias radiologists against the correct diagnosis, resulting in decreased diagnostic accuracy, increased case time, decreased confidence in the diagnosis provided, and increased mental effort. Participants were expected to be more confident in accurately diagnosed cases compared to misdiagnosed cases and less confident during miscued compared to correctly cued cases, whereas self-reported mental effort was predicted to be inversely related to diagnostic accuracy. We predicted that more-experienced radiologists would display improved diagnostic performance during correctly cued cases compared to less-experienced individuals (Manning et al., 2006; Manning, Gale, & Krupinski, 2005). However, because of expected heuristic biases and reliance on case indications (Blumenthal-Barby & Krieger, 2015), the more-experienced participants' performance advantage would be mitigated during cases that were miscued to the correct diagnosis. In terms of underlying processes, we predicted more-experienced participants to display increased saccadic amplitude and velocity (Kocak et al., 2005; Krupinski et al., 2006) and reduced time to first fixation on areas of interest (Wood et al., 2013), which would be associated with diagnostic accuracy on correctly cued cases. We refrain from providing directional hypotheses relating to whether more-experienced individuals would show increased fixation durations on areas of interest (i.e., reflecting task difficulty or individual sensitivity to abnormalities; Cooper et al., 2009; Evans, Cohen, et al., 2011) or decreased fixation durations (i.e., due to reduced viewing time needed to make a judgment; Bertram et al., 2013; Krupinski et al., 2006; Leong et al., 2007; Manning et al., 2006). Finally, based on previous work, we expected less-experienced participants would report fewer evaluative, prediction, and deep planning statements compared to the more-experienced group (McRobert et al., 2013; Roca et al., 2011), whereas all participants were expected to report less evaluative and deep planning thought processes during miscued compared to correctly cued cases (McRobert et al., 2011).

Method

Participants

Altogether, 21 radiologists were recruited through the radiology and imaging sciences department in the school of medicine at the lead institution by word of mouth (M_{age} = 34.26 years, SD = 5.80). Participants had completed medical school at various institutions across the country, with two participants completing their medical school training at the lead institution and one individual attending medical school outside of the U.S. All participants were accredited by the Liaison Committee on Medical Education with the exception of one participant who was a Doctor of Osteopathic Medicine. All read and signed an informed consent approved by the Institutional Review Board prior to starting the experiment and each received $50 compensation for participating.

Participants were divided into two groups of varying experience levels, with 11 radiology residents comprising the less-experienced group and 10 radiology fellows and attendings comprising the more-experienced group. To check the validity of the groupings, each radiologists' actual caseload/volume over the last 6 months were gathered from the radiology department and the university hospital. Moreover, a Career Practice History Questionnaire (PHQ), adapted from previous work (Ford, Low, McRobert, & Williams, 2010), was used to elicit participants' self-reported educational and training background. PHQ data were available for all participants except one resident who moved institutions. Questions pertained to the duration, proportion, and specificity of time spent engaged in active image interpretation within a variety of radiology subspecialties (e.g., abdominal, breast, cardiothoracic, neuroradiology, nuclear medicine, pediatrics, vascular, and musculoskeletal).

The less-experienced group had completed relatively fewer total cases in the previous six months (M = 537.09 ± 178.65) compared to the more-experienced group (M = 3,649.20 ± 3,490.25). The less-experienced group had completed a relatively smaller proportion of their caseload within the MSK subspecialty (M = 41.99% ± 24.73) compared to the more-experienced group (M = 67.75% ± 28.21). We expected greater variability in caseload in the more-experienced group, given that our sample ranged from 1st to 3rd year fellows to attending radiologists with over 10 years of experience postfellowship. The PHQ further revealed that more-experienced radiologists self-reported a greater proportion of their active image interpretation time spent within the MSK subspecialty across their entire career (M = 41.66% ± 19.04) compared to the less-experienced group (M = 17.60% ± 13.89). In sum, the more- and less-experienced groups differed in terms of their title, years of training, as well as their experience in diagnosing MSK cases.

Stimuli

Participants were presented with 16 deidentified musculoskeletal (MSK) radiographic cases provided by the Department of Radiology and Imaging Sciences. The cases contained between two to four separate radiographic images that could be viewed at the discretion of the radiologist, representing four distinct categories: present (positive) metastatic cancer; present trauma; present other; and absent (negative). Present cases included a primary abnormality within distinct, nonoverlapping locations (e.g., spine, shoulder, knee, hip anatomic areas), with no additional major findings present that could otherwise detract from the detection of the primary finding. The area(s) of interest was defined using an ellipse to demarcate the minimum size on each image necessary to cover the site(s) of abnormality. All abnormalities and their exact location on present/positive cases were confirmed on cross sectional (CT or MRI) or follow up radiographic studies. Absent/negative cases with no abnormal radiographic findings were presented on four cases. We included information on the patient's age, gender, and brief indication as to why the study was performed (i.e., trauma, evaluate hardware, history of cancer). Participants were positioned approximately 70 cm away from a research-dedicated, four-monitor system that mirrored the Phillips Picture Archiving and Communication System (PACS) workstation used in their radiology reading room to perform case dictation and diagnosis during a clinical shift. Participants could zoom in and out of each image, rotate each image, adjust the window width and level, and add annotations.

Context Manipulation

On four of the 12 present cases, the accompanying case notes (the clinical information provided with the case) were manipulated (i.e., mismatched from another case) to be miscued to the correct diagnosis, without changing information about the patient's sex or age (see Table 1). For example, a case in which the correct diagnosis was a "trauma, evaluate for compression fracture" was miscued to include a disingenuous indication of "history of cancer, experiencing back pain" (see Figure 1). Miscued case notes were used on two cases which were present for metastatic cancer (incorrectly cued as trauma cases) and two trauma cases (miscued as metastatic cancer). To avoid participants recognizing the manipulation, all miscued cases were randomly interspersed into the latter half of the case and case notes were never miscued on absent, present-other, or the remaining present cancer and present trauma cases.

Retrospective Verbal Reports

Participants were introduced to a method of providing retrospective think-aloud verbal reports adapted from Ericsson and Kirk (2001). During a short (~15 min) training session, participants were instructed to provide a retrospective think-aloud diary of the thoughts and cognitive processes used during the diagnosis. Each verbal report began with, "The first thought I had was . . ." and continued until participants described all the thoughts they had prior to completing the diagnosis. As opposed to the concurrent verbal reports used in previous work (e.g., McRobert et al., 2013, 2017), a retrospective approach was used to avoid interfering with the continuous verbal dictation of findings/impressions that occurs during daily practice. Participants were permitted to keep case notes and images open during the verbal report procedure to facilitate complete recollection but were discouraged from providing summaries of their thoughts or verbalizing new thoughts that occurred after the diagnosis was complete. Retrospective verbal reports were provided following the first two "practice" cases and again on two present-cued and two present-miscued cases selected at random from the latter half of the case sequence.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Case Pool With Description of the Correctly Cued Indication (Case Notes), Miscued Case Indication, and Correct Diagnosis

Case type	Case #	Correctly-cued indication	Miscued indication	Diagnosis
Present for trauma (pt)	pt-1	Trauma	History of cancer, back pain	Spine—L1 compression fracture
	pt-2	Acute injury, pain	Malignant neoplasm of the breast	Knee—proximal fibula fracture
	pt-3	Fall	Pain, renal cell carcinoma	Shoulder—greater tuberosity fracture
	pt-4	Ground level fall	Eval metastatic disease	Hip—sup and inf pubic rami fracture
Present for metastatic cancer (pm)	pm-1	Pain, eval for metastasis	Trauma	Spine—L3 breast met
	pm-2	Malignant neoplasm of the kidney	Injury	Knee—medial femoral renal cell carcinoma metastasis
	pm-3	Shoulder pain, history of breast cancer	Fall	Shoulder—acromial breast metastasis
	pm-4	Evaluate for metastatic disease	Motor vehicle accident	Hip—acetabular renal cell carcinoma metastasis
Present for other abnormality (po)	po-1	Pain, elevated ESR	—	Sacroilitis
	po-2	Eval hardware	—	Shoulder glenoid notching
	po-3	Follow up enchondroma	—	Knee enchondroma
	po-4	Chronic knee pain, evaluate osteoarthritis	—	Knee arthritis
	po-5	Eval hardware	—	Spine fusion complication
	po-6	Intermittent lower back pain	—	Spine disc degeneration
Absent cases (a)	a-1	Low back pain	—	Normal spine
	a-2	Post operation	—	Normal spine fusion
	a-3	Pain, no injury	—	Normal knee
	a-4	Pain and swelling	—	Normal shoulder
	a-5	Left hip pain	—	Normal hip
	a-6	Post operation	—	Normal THA

Note. sup = superior; inf = inferior; ESR = erythrocyte sedimentation rate; THA = total hip arthroplasty.

Procedure

Participants provided informed consent and completed the PHQ before the mobile eye tracking system was fitted and calibrated. Participants were instructed that they would be given a caseload containing a set of 16 musculoskeletal cases that they should dictate and provide impressions on as they would during normal clinical practice. They were informed that some cases contained a finding (i.e., positive for a major abnormality) and others did not contain any major finding (i.e., negative). The cases were given a unique identifier (e.g., alpha, echo, tango) to create six separate sequences that were pseudorandomized and counterbalanced within PACS. In each sequence, participants always dictated absent and present other cases that were correctly cued in the first five cases. This buffer period ensured that participants were not aware of the context manipulation. The first two cases were used as practice to establish familiarity with the protocol as well as the eye-tracking system, experimental procedures, and verbal reports. The remaining 11 cases included absent, present other, present metastatic cancer, and present trauma cases, with two present metastasis and two present trauma cases being incorrectly cued. Participants dictated each case as they normally would during daily practice without interruption and provided their self-reported ratings of confidence and mental effort after each case. The experimenters only stopped participants to complete retrospective verbal reports after two correctly and two miscued cases selected at random that were present for either trauma or cancer.

Measures

Diagnostic performance. The diagnostic reports were generated using Nuance PowerScribe 360 Reporting (Nuance, Burlington, MA), which is widely used by practicing physicians and trainees at the institution. The reports were coded for accuracy by two separate practicing radiologists on the research team. Accuracy was coded as binary (0 = inaccurate, 1 = accurate) with each aspect of the final report (e.g., findings, cause of abnormality, future recommendations) graded as either accurate or inaccurate and the final accuracy coded as accurate only if all aspects of the report were accurate. The typical sequence of events in each case was that the participant opened the case, images were displayed on the center two screens, participants pulled up the correct dictation report for that case and read the case notes, participants began viewing the images and finally, participants dictated their report. Participants would look back at the dictation screen periodically throughout the case to reread the case notes, reread their findings/impressions, and enter the correct field entry, or check/edit their spelling or grammar on their report. We quantified the total viewing time in each case (or case time) from the moment participants initially viewed the images, but only after reading the case notes, until the last fixation on the images before finishing their report.

Confidence and effort. Participants provided self-reported confidence in the diagnosis given on a scale of 1 (*uncertain*) to 5 (*certain*) and mental effort required to complete the case on a scale of 0 (markedly below average effort) to 5 (great amount of effort) after they finished dictating the report. These two Likert response formats were employed in order to provide a straightforward and expedient manner in which to assess case difficulty and confidence on each case. Previously, researchers have used Likert scales to measure confidence in diagnostic radiology (e.g., Ng & Palmer, 2007), but to our knowledge no researchers to date have assessed mental effort on a case-to-case basis. In an effort to expedite experimental time and avoid potential issues adapting other effort scales (e.g., Paas, Tuovinen, Tabbers, & Van Gerven, 2010; Zijl-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

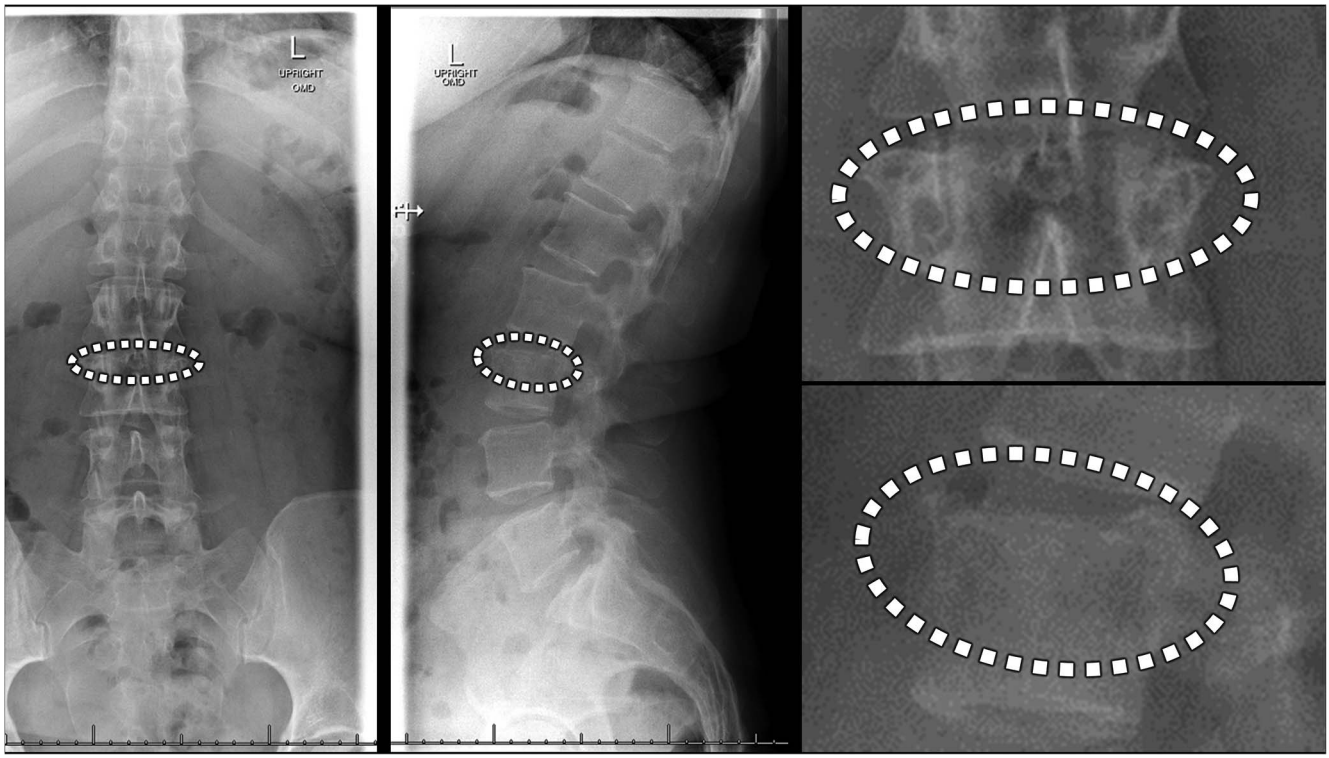

Figure 1. An example of a present metastatic case. The site of metastatic breast cancer of the L3 vertebrae is highlighted on both images for reference using dotted lines, and zoomed views of the areas of interest are provided to the right. When correctly cued, case notes read, "41 year-old female, pain, evaluate for mets." When miscued, case notes read, "41 year-old female, trauma."

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

stra & Van Doorn, 1985) to the PACS system, we transformed previous mental effort scales to a 6-point scale.

Gaze behavior. Visual search data were gathered using a lightweight, mobile head-mounted eye tracking system (ETG 2.0; Sensorimotor Instruments (SMI), Teltow, Germany) that sampled visual point of gaze at 120 Hz using a video-based monocular corneal reflection, accurate to within ±1° visual angle and within 1° in both vertical and horizontal fields with ±.5° precision. Visual fixations were determined based on established criteria (i.e., within 1° of visual angle for longer than 100 ms). Raw eye-tracking data (e.g., *x*-, *y*-coordinates over time) were uploaded into SMI BeGaze analysis software (v3.7), and gaze fixations were mapped onto key areas of interest for each case (e.g., the medical images, any abnormality on each image, the dictation/report screen). Fixations were quantified during post-processing using BeGaze™ semantic gaze mapping based on viewing area: interface; dictation screen; medical images; site of abnormality; other. Due to limits of precision for the eye tracking system, fixations that were within ± .5° of a viewing area (i.e., site of abnormality) were deemed to have occurred on that area. The number of fixations, number of fixations per second, mean fixation duration, and total time spent fixating on that area (i.e., dwell time) as a percent of case time were calculated. Saccades were identified automatically and subsequently quantified based on their spatiotemporal properties (e.g., mean saccadic amplitude, velocity, acceleration). Time to first fixation on the key area(s) of interest was calculated during cases present (positive) for an underlying abnormality.

Verbal report statements. Verbal reports were transcribed and segmented using natural speech patterns (i.e., syntactical markers) and subsequently categorically coded using a method adapted from Ericsson and colleagues (Ericsson & Kirk, 2001; Ericsson & Simon, 1993). Such an approach has been previously used with emergency room physicians (McRobert et al., 2013). Specifically, each report was coded based on whether it was a statement of cognition, evaluation, prediction, or deep planning. Coded categories were based on an established structure adapted from previous work (Ericsson & Simon, 1993; McRobert et al., 2011) in which cognitions were all statements representing current actions or recalled statements about current events (e.g., "I looked here") and evaluations included statements representing positive, neutral, or negative assessment. Predictions reflected statements about what would and could occur next, particularly in reference to the case history. Finally, deep planning was coded as statements representing a search for possible alternatives beyond the current diagnosis (e.g., "I would recommend an MRI"). The number of statements of each type was aggregated within each case and within each participant to compute the percentage of statement type as a function of the total number of statements made. Verbal reports were independently coded by three researchers and cross-analyzed for reliability using the intraobserver percentage agreement formula (Thomas & Nelson, 2001). The intraobserver agreement percentage for verbal report coding was 97.8%.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Statistical Analyses

Cases of each type (e.g., absent, present-cued, present-miscued) for each participant were aggregated to compute a mean score for each measure, including percent accuracy. In addition, each measure was averaged within a participant based on whether the case was diagnosed accurately versus inaccurately and whether it was cued or miscued. Any outliers (i.e., +/− 3 *SD* from the mean) for each dependent measure were removed prior to the analyses. Gaze data were removed for two participants due to poor tracking (i.e., fixation percent <60%). Statistical analyses are grouped into two sections: mixed-model analyses of variance (ANOVAs) to explore interactions between expertise and cue type; and linear mixed-effect regressions to measure changes in process-tracing measures as a function of diagnostic accuracy and cue type.

We conducted a 2 (group: experienced, less-experienced) × 2 (cue type: present-cued, present-miscued) mixed-model ANOVA with repeated measures on the last factor to identify differences in diagnostic performance as a function of experience and case context, including differences in gaze behavior on the area(s) of interest (i.e., site of the abnormality/lesion) and verbal report statements. Absent cases were analyzed separately with *t* tests to determine if differences (i.e., case time, false-positive rate) emerged as a function of experience. The assumption of normality supported across variables, based on Shapiro-Wilk tests. Significant main effects for each ANOVA were further decomposed using post hoc tests with Bonferroni-corrected Type I error rate.

To measure differences in diagnostic performance and gaze behavior for correctly and incorrectly diagnosed cases, we adopted linear-mixed effect regressions to allow for missing data within participants (e.g., some participants never correctly diagnosed a miscued case). These models had fixed-effects of diagnosis (accurate, inaccurate) and cue type (present-cued, present-miscued), with random-effects of participant crossed with these two factors. An inspection of the residuals from each model led us to log-transform the dependent variables, which resulted in a normal and homoscedastic distribution of residuals (Judd, McClelland, Ryan, McClelland, & Ryan, 2011). The Wald chi-squared test was used to assess the statistical significance of fixed-effects in these models. All inferential statistics are based on the log-transformed values. For ease of interpretation, however, all descriptive statistics are based on untransformed values.

The significance level for all statistical tests was set at $\alpha = .05$. Partial eta-squared (η^2) effect sizes were reported for ANOVA main effects and interactions. Cohen's *d* effect sizes (Cohen, 1992) were reported for pairwise comparisons in all statistical tests and were evaluated as trivial (0.00–0.19), small (0.20–0.49), medium (0.50–0.79), and large (0.80 and greater). Analyses were completed using SPSS 25.0 (Armonk, NY), R 3.4.1 (R Core Team, 2017), and R-Studio 1.1.463 (Boston, MA).

Results

Diagnostic Performance

Of the 120 of 261 cases that were misdiagnosed (i.e., false negative, false positive) across all cases, only four cases were classified as false positives. The more-experienced group misdiagnosed 47 cases, of which 12 (26.1%) were classified as perceptual errors, meaning that the participant did not visually fixate on the abnormality for longer than 1 s. The remaining 33 cases (69.6%) misdiagnosed by the more-experienced group were classified as decision-making errors. The less-experienced group misdiagnosed 73 cases, 15 (20.5%) of which were classified as perceptual errors, with the remaining 56 (76.7%) misdiagnosed cases classified as decision-making errors. No trends emerged in terms of perceptual or decision-making errors as a function of whether cases were cued or miscued (see Table 2).

Among the absent cases, only four false positives errors were made across groups. The only measure found to discriminate visual search behavior or performance on absent cases between groups was case time, $t(18) = 2.567$, $p = .019$, $d = 1.15$. The more-experienced participants completed absent cases faster ($M = 2.01$ min ± .84) than less-experienced participants ($M = 3.17$ min ± 1.16). Given the distribution of errors across experience groups, analyses focused on present cases.

Effects of experience and cue. There were significant main effects for group, $F(1, 19) = 7.634$, $p = .012$, $\eta^2 = .287$, and cue type, $F(1, 19) = 57.074$, $p < .001$, $\eta^2 = .750$, for diagnostic accuracy (Figure 2A), but no interaction for Group × Cue Type ($p > .05$). The more-experienced group ($M = 45.5\% \pm 25.6$) outperformed their less-experienced counterparts ($M = 29.4\% \pm 29.4$; $d = .58$). In addition, participants were significantly more accurate on present-cued ($M = 58.3\% \pm 20.5$) compared to present-miscued cases ($M = 17.5\% \pm 18.8$, $d = 2.08$). A significant Group × Cue Type interaction was documented for case time, $F(1, 18) = 7.762$, $p = .012$, $\eta^2 = .301$ (Figure 2B). Post hoc tests revealed that participants in the less-experienced group completed present-cued cases more slowly ($M = 3.61$ min ± 1.31) compared to present-miscued cases ($M = 2.90$ min ± 1.33; $p = .006$, $d = .62$).

A significant Group × Cue Type interaction was found for self-reported mental effort, $F(1, 19) = 5.175$, $p = .035$, $\eta^2 = .214$; however, post hoc tests comparing the effect of group and cue type did not reach significance (all *p*s > .05). A significant Group × Cue Type interaction was found for self-reported confidence in the diagnosis, $F(1, 38) = 3.60$, $p = .037$, $\eta^2 = .16$. Post hoc tests revealed that on present-cued cases more-experienced participants were more confident in their diagnosis ($M = 3.84 \pm .37$) compared to less-experienced participants ($M = 3.47 \pm .39$; $p = .04$, $d = .97$), and more-experienced participants reported higher confidence levels in their diagnosis on present-cued ($M = 3.84 \pm .37$) compared to present-miscued cases ($M = 3.31 \pm .59$; $p = .010$, $d = 1.08$).

Table 2
Types of False Negative Errors as a Function of Group and Cue Type

Cue type	Perceptual errors (<1 s fixation time)	Judgement errors (>1 s fixation time)	Total
Experienced			
Cued	8 (36.4%)	14 (63.6%)	22
Miscued	4 (17.4%)	19 (82.6%)	23
All cases	12 (26.7%)	33 (73.3%)	45
Less-experienced			
Cued	9 (25.0%)	27 (75.0%)	36
Miscued	6 (17.1%)	29 (82.9%)	35
All cases	15 (21.1%)	56 (78.9%)	71

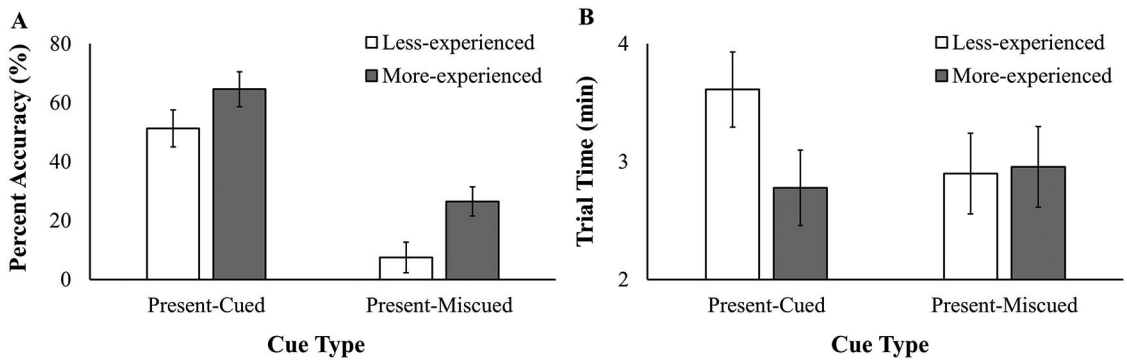

Figure 2. Mean and *SE* bars for (A) diagnostic accuracy percent and (B) case time in minutes as a function of group and cue type.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Effects of diagnostic accuracy and cue. For case time, the linear mixed-effect regressions revealed a significant main effect of diagnosis for case time, $\chi^2(1) = 13.23$, $p < .001$ (Figure 3A). Participants took longer on cases they diagnosed accurately (M = 3.50 min ± 1.35, $d = .65$) than those diagnosed inaccurately (M case time = 2.79 min ± 1.01).

For self-reported mental effort, there was a statistically significant Diagnosis × Cue Type interaction, $\chi^2(1) = 6.34$, $p = .012$. When participants were miscued, they rated correct cases as less effortful during inaccurately diagnosed cases (M effort score = 3.07 ± .89) than correctly diagnosed cases (M = 3.77 ±.93; p = .01, $d = .78$), whereas when they were correctly cued, they did not differ in their ratings on correctly and incorrectly diagnosed cases in terms of effort (M = 3.22 ± .59 and M = 3.10 ± .70; $p = .375$, $d = .18$). Furthermore, on cases that were correctly diagnosed, participants rated cases as requiring less mental effort when they were correctly cued (M = 3.10 ± .70) compared to when they were miscued (M = 3.77 ± .93, $p = .0459$, $d = .82$), but no differences in mental effort were documented during inaccurately diagnosed cases as a function of cue type ($p = .19$, $d = .20$).

For self-reported confidence, there were statistically significant main effects for diagnosis, $\chi^2(1) = 9.93$, $p = .002$, Case Type, $\chi^2(1) = 4.21$, $p = .040$, and a significant Diagnosis × Cue Type interaction, $\chi^2(1) = 8.63$, $p = .003$ (Figure 3B). When cases were miscued, participants reported lower confidence in their diagnosis on cases on which they were accurate (M confidence score = 2.95 ± 1.31) than on cases when an error was made (M = 3.53 ± 0.58; $p = .02$, $d = .57$). On correctly cued cases, however, participants tended to be more confident on cases where they made a correct diagnosis (M = 3.72 ± 0.44) than on cases where an error was made (M = 3.56 ± .57; $p = .19$, $d = .29$). On accurately diagnosed cases, participants were more confident when correctly cued (M = 3.71 ± .44) compared to when they were miscued (M = 2.95 ± 1.31, $p = .002$, $d = .78$), while no differences were documented on incorrectly diagnosed cases as a function of cue type. No other significant main effects or interactions were observed for diagnostic accuracy, case time, or self-reported mental effort and confidence (all ps > .05).

Visual Search Behavior

Effects of experience and cue. A significant main effect of cue type was reported for dwell time on the images as a percent of total case time (not including the area of interest), $F(1, 17)$ = 7.809, $p = .012$, $\eta^2 = .32$, revealing that participants spent less time viewing the medical images during present-cued cases (M dwell percent time = 66.6% ± 8.4) compared to present-miscued cases (M = 71.1% ± 9.5; $d = .50$). No other significant main effects or interactions were observed for visual search behavior (e.g., saccadic amplitude and velocity, fixations/s, fixation dura-

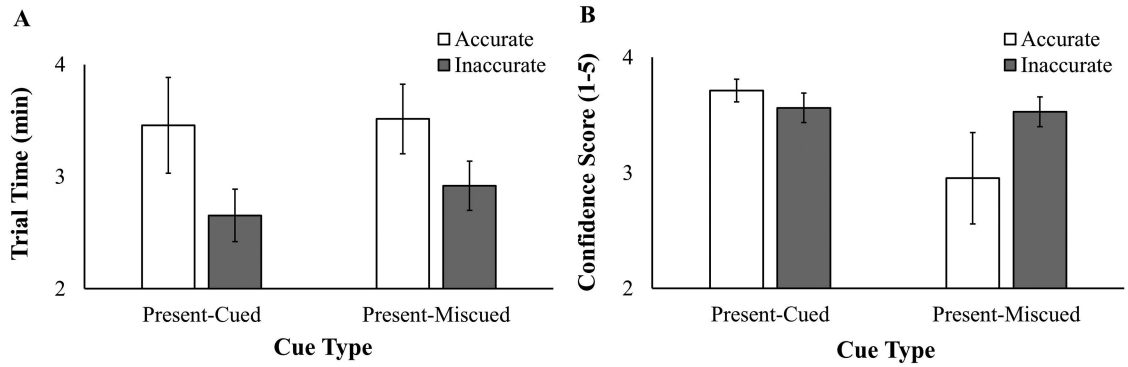

Figure 3. Mean and *SE* bars for (A) case time in minutes and (B) self-reported confidence as a function of diagnostic accuracy and cue type.

tion, dwell time) as a function of group or cue type (all $ps > .05$). In addition, no significant differences were documented based on case note quality in terms of the total number of fixations, fixations per second, mean fixation duration, or dwell time on key area(s) of interest where the abnormality was located (all $ps > .05$).

Effects of diagnostic accuracy and cue. Linear mixed-effect regressions revealed a significant Diagnosis × Cue Type interaction for average saccadic amplitude, $\chi^2(1) = 4.92$, $p = .026$, and a significant main effect of cue type for average saccadic velocity, $\chi^2(1) = 6.25$, $p = .012$ and peak velocity, $\chi^2(1) = 6.21$, $p = .013$, and average saccadic amplitude to first fixation on the area of interest, $\chi^2(1) = 4.832$, $p = .028$. Participants exhibited greater saccadic amplitude during cases they inaccurately diagnosed when they were correctly cued (M degrees = 5.97 ± 1.53) compared to cases that were miscued (M = 5.40 ± 1.69, $p = .01$, $d = .35$). Moreover, when participants were miscued, they exhibited greater saccadic amplitude when making accurate diagnoses (M = 6.05 ± 1.68) compared to inaccurately diagnosed cases (M = 5.40 ± 1.69, $p = .04$, $d = .11$; see Figure 4A). Saccades immediately preceding the first fixation on the area of interest were larger when participants were cued (M = 205.81 ± 30.8) compared to when they were miscued (M = 197.03 ± 36.86, $p = .046$, $d = .26$). The average (M degrees/s = 111.99 ± 34.17) and peak velocity (M degrees/s = 284.23 ± 171.97) of saccadic movements was greater during correctly cued cases compared to the average (M = 98.07 ± 16.93, $d = .52$) and peak velocity (M = 226.45 ± 4, $d = .44$) during miscued cases.

The mixed-effect regression analyses further revealed a significant main effect of diagnosis, $\chi^2(1) = 6.94$, $p = .008$, and case type, $\chi^2(1) = 5.18$, $p = .023$, for the total dwell time percent on the area(s) of interest (Figure 4B) and a diagnosis main effect for total time spent fixating on the area of interest, $\chi^2(1) = 12.218$, $p < .001$. Overall, participants spent a greater percentage of time viewing the area(s) of interest during accurately diagnosed cases (M dwell time percent = 8.08 ± 4.89) compared to inaccurately diagnosed cases (M = 5.51 ± 3.50, $d = .60$), as well as more total viewing time on the area(s) of interest during accurately (M seconds = 12.99 ± 8.41) compared to inaccurately diagnosed cases (M = 7.39 ± 5.04, $p = .001$, $d = .81$). Participants also spent a greater proportion of time viewing the area(s) of interest on cases that were miscued (M = 7.36 ± 4.23) compared to cases that were correctly cued (M = 6.06 ± 4.36, $d = .30$). In addition, significant main effect of diagnosis for the average fixation duration on the area(s) of interest, $\chi^2(1) = 25.31$, $p < .001$ (Figure 4C). Participants exhibited fixations of longer duration on the area(s) of interest during cases they accurately diagnosed (M time in ms = 357.58 ± 102.26) compared to those that were misdiagnosed (M = 268.59 ± 64.95, $d = .20$).

Verbal Reports

A significant main effect was reported for cue type for the proportion of statements made that were classified as prediction statements (i.e., statements about what would and could occur next based on previous information), $F(1, 19) = 4.40$, $p = .0497$, $\eta^2 = .188$. Follow-up tests indicated that participants self-reported making a greater proportion of prediction statements during present-cued cases (M = 19.84% ± 13.80) compared to present-miscued cases (M = 14.57% ± 10.98, $d = .42$). No other significant main effects or interactions were documented for the proportion of statement type made during the retrospective verbal report as a function of group or case type (all $ps > .05$).

Similarly, in linear mixed-models looking at the effect of diagnosis and cue type, there were no reliable differences in either the proportion of cognitive statements, evaluative statements, prediction statements, or planning statements. The frequency of these statements was not statistically different as a function of accuracy ($ps > .133$), cue ($ps > .053$), or the Diagnosis × Cue Type interaction ($ps > .190$).

Discussion

We examined whether changing context, by manipulating the accuracy of case notes, influenced perceptual-cognitive bias in more- and less-experienced radiographers. We predicted that miscued case notes would negatively bias radiologists against the correct diagnosis, resulting in decreased diagnostic accuracy, increased case time, decreased confidence in the diagnosis provided, and increased mental effort. Moreover, we expected that more-experienced radiologists would display improved diagnostic per-

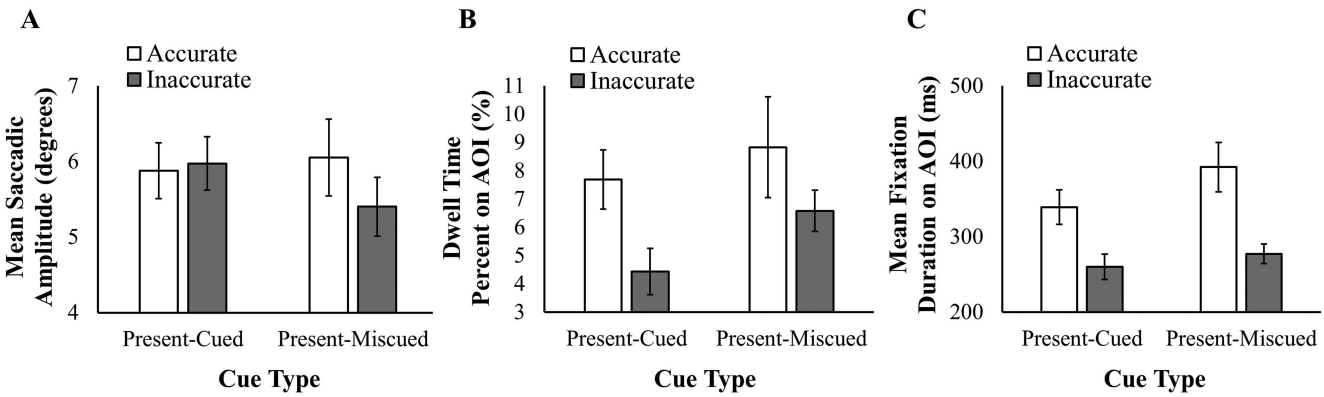

Figure 4. Mean and *SE* bars as a function of diagnostic accuracy and cue type for (A) mean saccadic amplitude in degrees, (B) mean dwell time percent on the area(s) of interest as a percent of total fixation time, and (C) mean fixation duration on the area(s) of interest in milliseconds.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

formance during correctly cued cases compared to less-experienced individuals, but that this performance advantage would be mitigated during cases that were miscued to the correct diagnosis. Finally, we expected more-experienced radiologists to exhibit increased efficiency of visual search (e.g., longer faster saccades) and more evaluative, prediction, and deep planning statements compared to less-experienced participants.

Several novel contributions emerged. First, miscued case notes increased perceptual-cognitive bias in both groups, resulting in a decrease of approximately 40% in diagnostic accuracy. Second, more-experienced radiologists outperformed less-experienced individuals, but no systematic differences in gaze behaviors emerged between groups. Third, most errors were judgment errors, meaning participants visually fixated on the abnormality for longer than a second, yet still failed to make the correct diagnosis. Finally, diagnostic accuracy could be predicted during correctly cued cases based on gaze behavior (e.g., longer time fixating on the area of interest, longer and faster saccades). We elaborate on these and other findings with reference to existing research on perceptual-cognitive expertise within diagnostic medicine and other domains.

We demonstrated that manipulating context through the provision of low-quality case notes was effective in increasing perceptual-cognitive bias in radiologists. Providing radiologists with case notes that were miscued to the correct diagnosis resulted in a 40% decrease in diagnostic accuracy. The low-quality or miscued case notes were not overtly deceptive (e.g., history of cancer, back pain for a trauma case), since many individuals in real clinical settings can be suffering from chronic conditions while concurrently experiencing acute injuries. Critically, our data suggest even a slight "miscue" can have a dramatic, negative effect on the ability to make accurate decisions regardless of experience and may prolong diagnostic decisions in less-experienced individuals. In clinical settings, cases where findings are retrospectively visible but are not initially detected are considered as "misses," regardless of what clinical information was provided at the time of interpretation. In malpractice lawsuits in which radiologists are held accountable for diagnostic errors, our findings suggest that some of those misses could be reasonably expected and may be within the expected standard of care, particularly if incorrect or misleading clinical information was present. For instance, if a radiologist is sued for missing a metastatic lesion in a patient who is being evaluated after falling down the stairs, should this be considered malpractice? These data suggest that a large proportion of radiologists would have missed the correct diagnosis of cancer if the patient's clinical information did not explicitly cue for it, or if the lesion was sufficiently difficult to locate. Perhaps the presence of poor or incorrect clinical information should be considered when judging if a diagnosis was truly a miss, reflecting the more measured approach advocated in a recent review of diagnostic errors (Brady, 2017). In the future, investigators should clarify these effects by examining perceptual-cognitive bias during decision making when radiology professionals do not have access to case notes/patient histories or are provided with even more explicit indications for the potential diagnosis.

Although the more-experienced radiographers displayed improved diagnostic accuracy, no systematic differences were documented in visual gaze behavior between groups. This finding was unexpected given the well-documented differences in gaze behavior that exist between more and less-experienced participants (Gegenfurtner et al., 2011). One important difference between our work and previous research in this area is that we used a mobile eye-tracker that allowed the radiologists to evaluate four separate screens, as they do in clinical practice. In contrast, almost all prior medical image perception research has used stationary eye-trackers that typically only allow one screen to be viewed. It is possible that some of the previously observed eye-tracking effects are a result of the relatively artificial viewing conditions, rather than a general finding that applies in clinical practice. Clearly, more research is needed to determine how viewing conditions interact with visual expertise.

The extant research on diagnostic imaging indicates that expert radiologists can detect abnormalities in medical images viewed for less than 300 ms (e.g., Carrigan, Wardle, & Rich, 2018; Evans, Haygood, Cooper, Culpan, & Wolfe, 2016; Kundel & Nodine, 1975; Mugglestone, Gale, Cowley, & Wilson, 1995), enough time for one saccade at most. In addition, *where* an individual fixates their gaze might not correspond to *what* information they extract from the visual scene (Ryu, Abernethy, Mann, Poolton, & Gorman, 2013), implying that inferences drawn solely from gaze behavior are prone to measurement bias. Because numerous factors can influence cognitive bias (Dawson & Arkes, 1987), and likely govern the use of gaze behaviors in clinical medicine, what constitutes effective and/or efficient performance during decision-making is largely dictated by task-specific demands and inter- and intra-individual differences in how performers use gaze to extract relevant information (see Mann, Causer, Nakamoto, & Runswick, 2019).

One potential interpretation of these findings is that individuals who have received at least a minimum amount of structured practice in a subspecialty do not tend to systematically view medical images differently from their peers. Alternatively, it could be reasonably argued that the lack of group differences in gaze behavior might be due to task difficulty, combined with interindividual variability in subspecialty and experience within both groups. For instance, some of the 3rd and 4th year residents in the sample were essentially 6–12 months away from being classified as "fellows" in terms of their objective training history, and their performance on the task reflected this experience. Likewise, some participants in the more-experienced group (including attending physicians with 10+ years of experience) may have lacked the necessary musculoskeletal subspecialty training, which helps to explain why their error rates and gaze behavior were more similar to those in the less-experienced group. In future work, researchers should seek to clarify some of these questions by tracking perceptual-cognitive expertise in diagnosticians across their residency training. A final interpretation of the lack of group differences is that systematic viewing practices, which have become commonplace in the domain (Kok et al., 2016), may not be as helpful when faced with incongruent information about the task (i.e., poor quality patient information). In these challenging situations, where an individual looks might be less important than what they are thinking about (Kok et al., 2012). More global or unique visual search strategies might therefore be beneficial to stave off perceptual-cognitive bias.

The absence of systematic group differences in visual search behavior does not imply that a superior search pattern cannot, and should not be modeled, for radiography trainees. In the present study, visual search strategies did appear to differentiate perfor-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

mance on cases that were accurately versus inaccurately diagnosed. Specifically, the total and percentage of time spent viewing key areas of interest was predictive of diagnostic accuracy, with individuals also displaying fewer fixations of longer duration on abnormalities during cases that were accurately diagnosed. Generally, accurately diagnosed cases were characterized by participants fixating on abnormalities for 5 s longer than cases that were inaccurately diagnosed. In addition, larger saccadic amplitudes were reported on cases that were accurately diagnosed when they were correctly cued compared to when they were miscued. This finding suggests that congruent contextual information guided visual search toward relevant information and away from irrelevant information in the display. However, miscued cases that were incorrectly diagnosed also tended to elicit longer saccades. These data highlight the potential pitfall of relying solely on gaze behavior to infer expertise during decision making and reinforce the need to use multiple-process tracing measures when examining the mechanisms underpinning superior performance (Williams & Ericsson, 2005).

In an attempt to address this latter concern, we used retrospective verbal reports to assess the cognitive processes and problem-solving skills used during the task. The proportion of predictive statements made differentiated cued versus miscued cases, reflecting that participants were either referring back to the case histories when describing what they expected to find (i.e., it was an older female and trauma, so I was alert for fractures), or using information gathered from the case to make subsequent judgments (i.e., "this patient is clearly osteopenic, which is concerning for nondisplaced fractures"). Overall, this finding supports the important role of accurate case notes in framing the context of a diagnostic evaluation.

Several published reports have demonstrated the utility of retrospective verbal reports to assess the cognitive processes underlying decision-making (e.g., McRobert et al., 2013; Roca, Ford, McRobert, & Williams, 2013). Our data were less illuminating than those in previous studies, but several limitations should be acknowledged. First, radiologists and diagnosticians typically speak throughout the diagnosis using assistive technology, making concurrent verbal reports not possible. Because of the retrospective design of our verbal report procedure, participants had to recall numerous thought processes and sequence those as they walked through the previous case aloud. For example, during the debriefing procedure, some participants self-reported having difficulty verbalizing all of the thought processes used during the case even after instruction and practice cases were provided to train this skill. Those that did verbalize numerous statements often only described the search sequence (e.g., "I looked here, then I looked here") rather than indicating why they chose particular search patterns; however, the degree to which verbal reports reflected these search sequences (i.e., cognitive statements) did not differ by experience or cue. Relatedly, there were few deep planning statements (e.g., follow-up procedures) made by radiologists in the verbal report procedure, perhaps reflecting that these radiologists did not know what to recommend moving forward given the difficulty of these cases. Although these data indicate that the quality of case notes may influence radiologists use of predictive strategies when diagnosing a case, questions remain regarding what other cognitive factors underlie superior decision-making in radiographic professionals and other pathologists. In particular, research is needed to refine verbal report coding procedures for diagnostic medicine domains like radiology that require complex cognitive processes to ascertain whether certain problem-solving skills are essential to most effectively extract and interpret relevant information correctly.

In conclusion, our findings suggest that even experienced radiologists are susceptible to perceptual-cognitive bias when the case notes or patient histories do not explicitly guide diagnostic decisions. Given that even experts may be unlikely to find rare or less prevalent abnormalities (Evans, Tambouret, et al., 2011; Wolfe et al., 2007), the quality of case notes should be of paramount importance to avoid incorrectly biasing a diagnostician to the wrong conclusion. Diagnostic professionals are practically and financially incentivized to quickly and accurately diagnose cases during daily practice, so these findings raise questions as to the acceptable margin for error in clinical practice, as well as significant concerns regarding how diagnosticians should use patient history information when making decisions about the potential presence or absence of certain abnormal findings. It could be reasonably argued that errors made using misleading or incomplete case notes should not be held against diagnosticians when considering malpractice in the field. In terms of the broader perceptual-cognitive literature, these findings support the influence of contextual bias on decision making in other domains (Dror, Kukucka, Kassin, & Zapf, 2018; Kassin, Dror, & Kukucka, 2013; Oliver, 2017), yet suggest that deficits in perception might not always be the root cause of faulty decision-making processes in medical professionals.

References

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. http://dx.doi.org/10.1016/j.chb.2016.09.022

Beam, C. A., Conant, E. F., Sickles, E. A., & Weinstein, S. P. (2003). Evaluation of proscriptive health care policy implementation in screening mammography. *Radiology, 229,* 534–540. http://dx.doi.org/10.1148/radiol.2292021585

Berlin, L. (2007). Radiologic errors and malpractice: A blurry distinction. *American Journal of Roentgenology*. American Journal of Roentgenology, *189,* 517–522. http://dx.doi.org/10.2214/AJR.07.2209

Bertram, R., Helle, L., Kaakinen, J. K., & Svedström, E. (2013). The effect of expertise on eye movement behaviour in medical image perception. *PLoS ONE, 8*(6), e66169. http://dx.doi.org/10.1371/journal.pone.0066169

Blumenthal-Barby, J. S., & Krieger, H. (2015). Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making, 35,* 539–557. http://dx.doi.org/10.1177/0272989X14547740

Bradley, J. D., Paulus, R., Komaki, R., Masters, G., Blumenschein, G., Schild, S., . . . Choy, H. (2015). Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): A randomised, two-by-two factorial phase 3 study. *The Lancet Oncology, 16,* 187–199. http://dx.doi.org/10.1016/S1470-2045(14)71207-0

Brady, A. P. (2017). Error and discrepancy in radiology: Inevitable or avoidable? *Insights Into Imaging, 8,* 171–182. http://dx.doi.org/10.1007/s13244-016-0534-1

Brady, A., Laoide, R. Ó., McCarthy, P., & McDermott, R. (2012). Discrepancy and error in radiology: Concepts, causes and consequences. *The Ulster Medical Journal, 81,* 3–9. Retrieved from www.ums.ac.uk

Bruno, M. A., Walker, E. A., & Abujudeh, H. H. (2015). Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction. *Radiographics, 35,* 1668–1676. http://dx.doi.org/10.1148/rg.2015150023

Carrigan, A. J., Wardle, S. G., & Rich, A. N. (2018). Finding cancer in mammograms: If you know it's there, do you know where? *Cognitive Research: Principles and Implications, 3,* 10. http://dx.doi.org/10.1186/s41235-018-0096-5

Causer, J., Barach, P., & Williams, A. M. (2014). Expertise in medicine: Using the expert performance approach to improve simulation training. *Medical Education, 48,* 115–123. http://dx.doi.org/10.1111/medu.12306

Choudhry, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). Systematic review: The relationship between clinical experience and quality of health care. *Annals of Internal Medicine, 142,* 260–273. http://dx.doi.org/10.7326/0003-4819-142-4-200502150-00008

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Cooper, L., Gale, A., Darker, I., Toms, A., & Saada, J. (2009). Radiology image perception and observer performance: How does expertise and clinical information alter interpretation? Stroke detection explored through eye-tracking. In B. Sahiner & D. J. Manning (Eds.), *Medical imaging 2009: Image perception, observer performance, and technology assessment* (Vol. 7263, pp. 72630K–72630K-12). Bellingham, Washington: SPIE.

Cooper, L., Gale, A., Saada, J., Gedela, S., Scott, H., & Toms, A. (2010). The assessment of stroke multidimensional CT and MR imaging using eye movement analysis: Does modality preference enhance observer performance? In D. J. Manning & C. K. Abbey (Eds.), *Medical imaging: Image perception, observer performance, and technology assessment* (Vol. 7627, pp. 76270B–76270B-12). Bellingham, WA: SPIE.

Cormier, E. M., Pickett-Hauber, R., & Whyte, J., IV. (2010). Cognitions and clinical performance: A comparison of high and low performing baccalaureate nursing students. *International Journal of Nursing Education Scholarship, 7*(1), e27. http://dx.doi.org/10.2202/1548-923X.2045

Dawson, N. V., & Arkes, H. R. (1987). Systematic errors in medical decision making: Judgment limitations. *Journal of General Internal Medicine, 2,* 183–187. http://dx.doi.org/10.1007/BF02596149

Donovan, T., & Manning, D. J. (2006). Successful reporting by non-medical practitioners such as radiographers, will always be task-specific and limited in scope. *Radiology, 12,* 7–12. http://dx.doi.org/10.1016/j.radi.2005.01.004

Drew, T., Võ, M. L. H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological Science, 24,* 1848–1853. http://dx.doi.org/10.1177/0956797613479386

Dror, I. E., Kukucka, J., Kassin, S. M., & Zapf, P. A. (2018). When expert decision making goes wrong: Consensus, bias, the role of experts, and accuracy. *Journal of Applied Research in Memory & Cognition, 7,* 162–163. http://dx.doi.org/10.1016/j.jarmac.2018.01.007

Egglin, T. K. P., & Feinstein, A. R. (1996). Context bias: A problem in diagnostic radiology. *Journal of the American Medical Association, 276,* 1752–1755. http://dx.doi.org/10.1001/jama.1996.03540210060035

Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: A perspective from the expert-performance approach with deliberate practice. *Academic Medicine, 90,* 1471–1486. http://dx.doi.org/10.1097/ACM.0000000000000939

Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (Eds.). (2018). *The Cambridge handbook of expertise and expert performance* (2nd ed.). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Kirk, W. (2001). *Instructions for giving retrospective verbal reports*. Unpublished manuscript, Florida State University, Tallahassee, FL.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47,* 273–305. http://dx.doi.org/10.1146/annurev.psych.47.1.273

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: Bradford Books.

Evans, K. K., Cohen, M. A., Tambouret, R., Horowitz, T., Kreindel, E., & Wolfe, J. M. (2011). Does visual expertise improve visual recognition memory? *Attention, Perception, & Psychophysics, 73,* 30–35. http://dx.doi.org/10.3758/s13414-010-0022-5

Evans, K. K., Haygood, T. M., Cooper, J., Culpan, A.-M., & Wolfe, J. M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 10292–10297. http://dx.doi.org/10.1073/pnas.1606187113

Evans, K. K., Tambouret, R. H., Evered, A., Wilbur, D. C., & Wolfe, J. M. (2011). Prevalence of abnormalities influences cytologists' error rates in screening for cervical cancer. *Archives of Pathology & Laboratory Medicine, 135,* 1557–1560. http://dx.doi.org/10.5858/arpa.2010-0739-OA

Ford, P. R., Low, J., McRobert, A. P., & Williams, A. M. (2010). Developmental activities that contribute to high or low performance by elite cricket batters when recognizing type of delivery from bowlers' advanced postural cues. *Journal of Sport & Exercise Psychology, 32,* 638–654. http://dx.doi.org/10.1123/jsep.32.5.638

Garland, L. H. (1949). On the scientific evaluation of diagnostic procedures. *Radiology, 52,* 309–328. http://dx.doi.org/10.1148/52.3.309

Gegenfurtner, A., Kok, E., van Geel, K., de Bruin, A., Jarodzka, H., Szulewski, A., & van Merriënboer, J. J. G. (2017). The challenges of studying visual expertise in medical image diagnosis. *Medical Education, 51,* 97–104. http://dx.doi.org/10.1111/medu.13205

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,* 523–552. http://dx.doi.org/10.1007/s10648-011-9174-7

George, J. E., Espinosa, J. A., & Quattrone, M. S. (1992). Legal issues in emergency radiology. Practical strategies to reduce risk. *Emergency Medicine Clinics of North America, 10,* 179–203.

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, 54,* 382–391. http://dx.doi.org/10.1053/j.jfas.2014.08.013

Graber, M. L., Franklin, N., & Gordon, R. R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine, 165,* 1493–1499. http://dx.doi.org/10.1001/archinte.165.13.1493

Grünheid, T., Hollevoet, D. A., Miller, J. R., & Larson, B. E. (2013). Visual scan behavior of new and experienced clinicians assessing panoramic radiographs. *Journal of the World Federation of Orthodontists, 2*(1), e3–e7. http://dx.doi.org/10.1016/j.ejwf.2012.12.002

Hall, K. H. (2002). Reviewing intuitive decision-making and uncertainty: The implications for medical education. *Medical Education, 36,* 216–224. http://dx.doi.org/10.1046/j.1365-2923.2002.01140.x

Judd, C. M., McClelland, G. H., Ryan, C. S., McClelland, G. H., & Ryan, C. S. (2011). *Data analysis*. New York, NY: Routledge. http://dx.doi.org/10.4324/9780203892053

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied*

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Research in Memory & Cognition, 2, 42–52. http://dx.doi.org/10.1016/j.jarmac.2013.01.001

Kelly, B. S., Rainford, L. A., Darcy, S. P., Kavanagh, E. C., & Toomey, R. J. (2016). The development of expertise in radiology: in chest radiograph interpretation, "expert" search pattern may predate "expert" levels of diagnostic accuracy for pneumothorax identification. *Radiology, 280,* 252–260. http://dx.doi.org/10.1148/radiol.2016150409

Kocak, E., Ober, J., Berme, N., & Melvin, W. S. (2005). Eye motion parameters correlate with level of experience in video-assisted surgery: Objective testing of three tasks. *Journal of Laparoendoscopic & Advanced Surgical Techniques Part A, 15,* 575–580. http://dx.doi.org/10.1089/lap.2005.15.575

Kok, E. M., De Bruin, A. B. H., Robben, S. G. F., & Van Merriënboer, J. J. G. (2012). Looking in the same manner but seeing it differently: Bottom-up and expertise effects in radiology. *Applied Cognitive Psychology, 25,* 854–862. http://dx.doi.org/10.1002/acp.2886

Kok, E. M., Jarodzka, H., de Bruin, A. B. H., BinAmir, H. A., Robben, S. G., & van Merriënboer, J. J. (2016). Systematic viewing in radiology: Seeing more, missing less? *Advances in Health Sciences Education, 21,* 189–205. http://dx.doi.org/10.1007/s10459-015-9624-y

Krupinski, E. A. (1996). Visual scanning patterns of radiologists searching mammograms. *Academic Radiology, 3,* 137–144. http://dx.doi.org/10.1016/S1076-6332(05)80381-2

Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics, 72,* 1205–1217. http://dx.doi.org/10.3758/APP.72.5.1205

Krupinski, E. A. (2011). The role of perception in imaging: Past and future. *Seminars in Nuclear Medicine, 41,* 392–400. http://dx.doi.org/10.1053/j.semnuclmed.2011.05.002

Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. T., 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,* 1543–1556. http://dx.doi.org/10.1016/j.humpath.2006.08.024

Kundel, H. L., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116,* 527–532. http://dx.doi.org/10.1148/116.3.527

Lee, C. S., Nagy, P. G., Weaver, S. J., & Newman-Toker, D. E. (2013). Cognitive and system factors contributing to diagnostic errors in radiology. *American Journal of Roentgenology*. American Journal of Roentgenology, *201,* 611–617. http://dx.doi.org/10.2214/AJR.12.10375

Leong, J. J. H., Nicolaou, M., Emery, R. J., Darzi, A. W., & Yang, G.-Z. (2007). Visual search behaviour in skeletal radiographs: A cross-specialty study. *Clinical Radiology, 62,* 1069–1077. http://dx.doi.org/10.1016/j.crad.2007.05.008

Loy, C. T. (2004). Accuracy of diagnostic tests read a systematic review. *The Journal of American Medical Assocation, 292,* 1602–1609.

Makary, M. A., & Daniel, M. (2016). Medical error-the third leading cause of death in the U.S. *British Medical Journal, 353,* i2139. http://dx.doi.org/10.1136/bmj.i2139

Mallett, S. (2014). Tracking eye gaze during interpretation of three-dimensional CT colonography materials and methods video data and readers. *Radiology, 273,* 783–792. http://dx.doi.org/10.1148/radiol.14132896

Mann, D. L., Causer, J., Nakamoto, H., & Runswick, O. R. (2019). Visual search behaviours in expert perceptual judgements. In *Anticipation and Decision-Making in Sport* (pp. 59–78). New York, NY: Routledge. http://dx.doi.org/10.4324/9781315146270-4

Manning, D. J., 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,* 134–142. http://dx.doi.org/10.1016/j.radi.2005.02.003

Manning, D. J., Gale, A., & Krupinski, E. A. (2005). Perception research in medical imaging. *The British Journal of Radiology, 78,* 683–685. http://dx.doi.org/10.1259/bjr/72087985

Marewski, J. N., & Gigerenzer, G. (2012). Heuristic decision making in medicine. *Dialogues in Clinical Neuroscience, 14,* 77–89.

McRobert, A. P., Causer, J., Vassiliadis, J., Watterson, L., Kwan, J., & Williams, M. A. (2013). Contextual information influences diagnosis accuracy and decision making in simulated emergency medicine emergencies. *British Medical Journal Quality & Safety, 22,* 478–484. http://dx.doi.org/10.1136/bmjqs-2012-000972

McRobert, A. P., Mercer, S. J., Raw, D., Goulding, J., & Williams, A. M. (2017). Effect of expertise on diagnosis accuracy, non-technical skills and thought processes during simulated high-fidelity anaesthetist scenarios. *British Medical Journal Simulation & Technology Enhanced Learning, 3,* 43–47. http://dx.doi.org/10.1136/bmjstel-2016-000129

McRobert, A. P., Ward, P., Eccles, D. W., & Williams, A. M. (2011). The effect of manipulating context-specific information on perceptual-cognitive processes during a simulated anticipation task. *British Journal of Psychology, 102,* 519–534. http://dx.doi.org/10.1111/j.2044-8295.2010.02013.x

Mugglestone, M. D., Gale, A. G., Cowley, H. C., & Wilson, A. R. M. (1995). Diagnostic performance on briefly presented mammographic images. In H. L. Kundel (Ed.), *Medical imaging 1995: Image perception* (Vol. 2436, pp. 106–116). Bellingham, WA: SPIE. http://dx.doi.org/10.1117/12.206840

Ng, C. S., & Palmer, C. R. (2007). Analysis of diagnostic confidence and diagnostic accuracy: A unified framework. *The British Journal of Radiology, 80,* 152–160. http://dx.doi.org/10.1259/bjr/64096611

Nodine, C. F., Mello-Thoms, C., Kundel, H. L., & Weinstein, S. P. (2002). Time course of perception and decision making during mammographic interpretation. *American Journal of Roentgenology, 179,* 917–923. http://dx.doi.org/10.2214/ajr.179.4.1790917

Oliver, W. R. (2017). Effect of history and context on forensic pathologist interpretation of photographs of patterned injury of the skin. *Journal of Forensic Sciences, 62,* 1500–1505. http://dx.doi.org/10.1111/1556-4029.13449

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2010). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38,* 63–71.

Pfeffer, M. R., Rabin, T., Tsvang, L., Goffman, J., Rosen, N., & Symon, Z. (2004). Orbital lymphoma: Is it necessary to treat the entire orbit? *International Journal of Radiation Oncology, Biology, Physics, 60,* 527–530. http://dx.doi.org/10.1016/j.ijrobp.2004.03.039

Roca, A., Ford, P. R., McRobert, A. P., & Williams, A. M. (2011). Identifying the processes underpinning anticipation and decision-making in a dynamic time-constrained task. *Cognitive Processing, 12,* 301–310. http://dx.doi.org/10.1007/s10339-011-0392-1

Roca, A., Ford, P. R., McRobert, A. P., & Williams, A. M. (2013). Perceptual-cognitive skills and their interaction as a function of task constraints in soccer. *Journal of Sport & Exercise Psychology, 35,* 144–155. http://dx.doi.org/10.1123/jsep.35.2.144

Rosenkrantz, A. B., & Bansal, N. K. (2016). Diagnostic errors in abdominopelvic CT interpretation: Characterization based on report addenda. *Abdominal Radiology, 41,* 1793–1799. http://dx.doi.org/10.1007/s00261-016-0741-8

Runswick, O. R., Roca, A., Williams, A. M., Bezodis, N. E., & North, J. S. (2018). The effects of anxiety and situation-specific context on perceptual–motor skill: A multi-level investigation. *Psychological Research, 82,* 708–719.

Ryu, D., Abernethy, B., Mann, D. L., Poolton, J. M., & Gorman, A. D. (2013). The role of central and peripheral vision in expert decision making. *Perception, 42,* 591–607. http://dx.doi.org/10.1068/p7487

Saber Tehrani, A. S., Lee, H., Mathews, S. C., Shore, A., Makary, M. A., Pronovost, P. J., & Newman-Toker, D. E. (2013). 25-year summary of

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

US malpractice claims for diagnostic errors 1986–2010: An analysis from the National Practitioner Data. *BMJ Quality & Safety, 22,* 672–680. http://dx.doi.org/10.1136/bmjqs-2012-001550

Shojania, K. G., & Dixon-Woods, M. (2017). Estimating deaths due to medical error: The ongoing controversy and why it matters. *British Medical Journal Quality & Safety, 26,* 423–428.

Siewert, B., Brook, O. R., Hochman, M., & Eisenberg, R. L. (2016). Impact of communication errors in radiology on patient care, customer satisfaction, and work-flow efficiency. *American Journal of Roentgenology, 206,* 573–579. http://dx.doi.org/10.2214/AJR.15.15117

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28,* 1059–1074. http://dx.doi.org/10.1068/p281059

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13,* 403–409. http://dx.doi.org/10.1016/j.tics.2009.06.003

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

Thomas, J. R., & Nelson, J. K. (2001). *Research methods in physical activity*. Champaign, IL: Human Kinetics.

van der Gijp, A., Ravesloot, C. J., Jarodzka, H., van der Schaaf, M. F., van der Schaaf, I. C., van Schaik, 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: Theory and Practice, 22,* 765–787. http://dx.doi.org/10.1007/s10459-016-9698-1

Wegwarth, O., Gaissmaier, W., & Gigerenzer, G. (2009). Smart strategies for doctors and doctors-in-training: Heuristics in medicine. *Medical Education, 43,* 721–728. http://dx.doi.org/10.1111/j.1365-2923.2009.03359.x

Whyte, J., IV, Cormier, E., & Pickett-Hauber, R. (2010). Cognitions associated with nurse performance: A comparison of concurrent and retrospective verbal reports of nurse performance in a simulated task environment. *International Journal of Nursing Studies, 47,* 446–451. http://dx.doi.org/10.1016/j.ijnurstu.2009.09.001

Williams, A. M., & Ericsson, K. A. (2005). Perceptual-cognitive expertise in sport: Some considerations when applying the expert performance approach. *Human Movement Science, 24,* 283–307. http://dx.doi.org/10.1016/j.humov.2005.06.002

Wolfe, J. M., Evans, K. K., Drew, T., Aizenman, A., & Josephs, E. (2016). How do radiologists use the human search engine? *Radiation Protection Dosimetry, 169,* 24–31. http://dx.doi.org/10.1093/rpd/ncv501

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General, 136,* 623–638. http://dx.doi.org/10.1037/0096-3445.136.4.623

Wood, G., Knapp, K. M., Rock, B., Cousens, C., Roobottom, C., & Wilson, M. R. (2013). Visual expertise in detecting and diagnosing skeletal fractures. *Skeletal Radiology, 42,* 165–172. http://dx.doi.org/10.1007/s00256-012-1503-5

Zijlstra, F. R. H., & Van Doorn, L. (1985). *The construction of a scale to measure perceived effort* (Technical report). Delft, the Netherlands: Delft University of Technology.

Received July 2, 2019
Revision received January 22, 2020
Accepted February 22, 2020 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.