Characteristics of expert search behavior in volumetric medical image interpretation

Lauren H. Williams[a,*] **Ann J. Carrigan**[b,c,d] **Megan Mills**[e] **William F. Auffermann**[e] **Anina N. Rich**[c,d,f] **and Trafton Drew**[g]

[a]University of California, San Diego, Department of Psychology, San Diego, California, United States

[b]Macquarie University, Department of Psychology, Sydney, New South Wales, Australia

[c]Macquarie University, Perception in Action Research Centre, Sydney, New South Wales, Australia

[d]Macquarie University, Centre for Elite Performance, Expertise, and Training, Sydney, New South Wales, Australia

[e]University of Utah, School of Medicine, Department of Radiology and Imaging Sciences, Salt Lake City, Utah, United States

[f]Macquarie University, Department of Cognitive Science, Sydney, New South Wales, Australia

[g]University of Utah, Department of Psychology, Salt Lake City, Utah, United States

Abstract

Purpose: Experienced radiologists have enhanced global processing ability relative to novices, allowing experts to rapidly detect medical abnormalities without performing an exhaustive search. However, evidence for global processing models is primarily limited to two-dimensional image interpretation, and it is unclear whether these findings generalize to volumetric images, which are widely used in clinical practice. We examined whether radiologists searching volumetric images use methods consistent with global processing models of expertise. In addition, we investigated whether search strategy (scanning/drilling) differs with experience level.

Approach: Fifty radiologists with a wide range of experience evaluated chest computed-tomography scans for lung nodules while their eye movements and scrolling behaviors were tracked. Multiple linear regressions were used to determine: (1) how search behaviors differed with years of experience and the number of chest CTs evaluated per week and (2) which search behaviors predicted better performance.

Results: Contrary to global processing models based on 2D images, experience was unrelated to measures of global processing (saccadic amplitude, coverage, time to first fixation, search time, and depth passes) in this task. Drilling behavior was associated with better accuracy than scanning behavior when controlling for observer experience. Greater image coverage was a strong predictor of task accuracy.

Conclusions: Global processing ability may play a relatively small role in volumetric image interpretation, where global scene statistics are not available to radiologists in a single glance. Rather, in volumetric images, it may be more important to engage in search strategies that support a more thorough search of the image.

 [DOI: 10.1117/1.JMI.8.4.041208]

Keywords: medical image perception; gist processing; expertise; scanners and drillers; lung cancer detection.

Paper 20348SSRR received Dec. 22, 2020; accepted for publication Jun. 28, 2021; published online Jul. 14, 2021.

*Address all correspondence to Lauren H. Williams, l8williams@ucsd.edu

2329-4302/2021/$28.00

1 Introduction

Identifying an abnormality in a medical image is a critical step toward patient diagnosis and treatment. However, medical image interpretation is a difficult task, and research spanning the past several decades has consistently revealed missed abnormality rates of ∼30%.[1] Given the challenge of this task, one might expect abnormality detection to involve an exhaustive search of the image until an abnormality is located. However, radiologists frequently report sensing an abnormality is present before it is actually located and identified in the image. Consistent with these anecdotal reports, radiologists detect most abnormalities within the first second of interpretation, which is much less time than it would take to complete an exhaustive search of the image.[2–4] In addition, radiologists can discriminate between normal and cancerous cases at a rate well-above chance after viewing medical images for only a fraction of a second.[5–8] These findings demonstrate that radiologists can extract a remarkable amount of information about a medical image in only a single glance. This phenomenon is referred to as "gist" or "global" processing, and these enhanced perceptual abilities are considered to be a key distinguishing characteristic between experts and novices in radiology.[9–11]

Although radiologists would never view medical images for only a fraction of a second in clinical practice—accuracy greatly improves with an unlimited viewing time[12]—these findings provide important insight on how the development of perceptual expertise influences naturalistic search behavior in radiology. Since the early 1970s, researchers have observed both qualitative and quantitative differences in search patterns across radiologists with different levels of experience.[13] More experienced radiologists have lower image coverage, make fewer fixations, have larger saccadic amplitude, and fixate on abnormalities more quickly (i.e., shorter time to first fixation) than both naïve and novice observers.[10] These findings suggest that experienced radiologists are able to rely more on the global properties of the image for attentional guidance than novices. These enhanced perceptual abilities appear to emerge before expert decision-making abilities develop and without any explicit instruction on search strategy.[13,14]

The differences in search behavior between experts and novices have led to a number of medical image perception models, each of which posits a major role for global processing in medical image interpretation.[15–17] The most recent of these models proposes a two-component visual search process with a non-selective (global) pathway and a selective (local) pathway that operate in parallel.[15] The non-selective pathway enables radiologists to rapidly extract the global statistical properties of an image. Although the non-selective pathway helps guide attention to perturbations in the image, detailed information about the abnormalities appears to be limited relative to the selective pathway.[6,7] In contrast, the selective pathway is limited in processing capacity but provides fine-grained information that supports the recognition and localization of abnormalities during a more foveal search. This two-pathway model originates in the visual search literature,[18] where evidence suggests that global summary statistics (e.g., mean size[19] and orientation[20] of objects, scene category,[21] or direction of motion[22]) can be extracted from scenes in a single glance, whereas only a limited number of objects can be recognized simultaneously due to limits of object-based attention.[23] Global processing ability in radiology is thought to involve the same cognitive mechanisms that allow laypeople to categorize familiar types of scenes after brief image presentations.[15,24,25] Through experience, radiologists develop a strong mental representation of a normal medical image, resulting in greater sensitivity to the statistical irregularities associated with an abnormal image. Thus more experienced radiologists are able to rely more on the non-selective pathway than novices, resulting in a search that relies more on information extracted from the periphery than an exhaustive search of the image.

Despite the prominent role of global processing in all major medical image perception models, some caution is warranted on the generalizability of these findings. These models were established using a relatively limited set of tasks: lung cancer detection using chest radiographs and breast cancer detection in mammography. Meanwhile, advancements in medical imaging technology have dramatically changed the size and complexity of medical images over the past several decades. In particular, there has been a shift from two-dimensional (2D) medical images, such as radiographs, to volumetric images, such as computed tomography (CT) scans, that better preserve the underlying three-dimensional (3D) structure of the human body. Volumetric medical images make up an increasingly large portion of radiologists' workload,[26,27] but it remains

unclear how global processing ability might manifest in these images, where the global statistical information is embedded in a navigable volume rather than being available to the observer in a single glance.[28]

Recent studies have evaluated global processing in these new modalities by showing observers videos of volumetric medical images that rapidly transition through the image slices at a fixed rate.[29,30] In these studies, observers were able to reliably discriminate between normal and abnormal cases after rapid image presentations and discrimination ability increased with observer experience. Although these studies provide evidence that global processing may play a role in volumetric image interpretation, we do not yet know how experience influences naturalistic search behavior. If more experienced radiologists use a global search strategy, eye tracking metrics associated with experience in 2D medical images, such as reduced image coverage and shorter time to first fixation, should replicate in volumetric image interpretation tasks. However, a recent review paper found that very few of these expertise-related differences in search behavior have been examined in comparable tasks using volumetric images.[28]

In addition to differences in scan patterns, global processing ability might also change how the observer scrolls through the depth of volumetric images. The global statistical properties of volumetric images are embedded throughout multiple stacked slices. Therefore, forming a global impression must involve some type of interaction with scrolling behavior. For example, an observer might establish a global impression of the image by frequently scrolling through the full depth of the image volume.[31] In a recent longitudinal study, radiology residents spent less time conducting "full runs" through the stack toward the end of their training, suggesting that global impressions of the image are established more efficiently with experience.[32] Similarly, experts adapted to faster image presentation speeds more easily than novices, which might reflect a shift toward a more global search strategy with experience.[33] However, other studies have not found any differences in performance between experts and novices at different image presentation speeds, and very few studies have addressed this question while allowing radiologists to freely scroll through the image stack as they would in clinical practice.[34,35]

Although global processing ability explains much of the variation between experts and novices in 2D image interpretation tasks, volumetric images introduce other aspects of search behavior that may help explain individual differences in performance. For example, two different strategies have been identified for searching through the depth of chest CT stacks during a lung cancer detection task: scanning and drilling.[36] Scanners search broadly across each slice of the CT scan while slowly moving through the image slices. In contrast, drillers keep their eyes relatively fixed in a single region of the lung at a time while rapidly scrolling through the depth of the stack. When given a fixed time limit for each case (3 min), drillers detected more lung nodules and had greater image coverage than scanners. These differences in performance are attributed to the fact that lung cancer nodules appear to flicker in and out of view when the observer scrolls through the image slices, which helps the observer differentiate the nodules from other structures, such as blood vessels, that persist throughout many slices of the image.[37]

It is not yet clear if the benefits of drilling generalize to tasks beyond lung cancer detection.[38,39] However, volumetric images clearly have unique properties that are important to consider in models of perceptual expertise. For example, lung nodules may appear to flicker in and out of view as the observer scrolls through the depth of a CT stack, which may mimic abrupt motion onset cues that are thought to involuntarily capture attention.[40] Although there does not appear to be a standard practice for how to instruct radiologists to search through volumetric medical images, search strategy might develop organically with experience. For example, a wider useful field of view (UFOV) might allow more experienced radiologists to take advantage of motion onset cues elicited in the periphery when scrolling through depth. Alternatively, search strategy might be passed on informally from mentor to mentee during training, or radiologists may simply learn that one strategy is more effective than another and begin to adopt it over time. In the original scanner/driller study, drillers reported reading more CT images in an average week than scanners, but radiologists in each group had similar years of experience.[36] Although this preliminary evidence that drillers had more regular experience with CT images is promising, that study was not designed to look at experience-related effects on search strategy, requiring more work to fully disentangle the effects of experience versus search strategy on task performance.

In sum, knowledge of how expert search behavior develops in volumetric image interpretation is currently a substantial gap in the medical image perception literature. Here we sought to help fill this gap by characterizing expert search behavior in a large sample of radiologists ($n = 50$) with a wide range of experience. In this study, radiologists evaluated chest CT scans for lung cancer nodules. Because lung cancer detection is one of the most well-researched tasks in the medical image perception literature, these findings can be more easily compared to the previous research. The first aim of this study was to determine whether behavioral and eye tracking measures associated with global processing ability in 2D images (accuracy, search time, image coverage, saccadic amplitude, and time to first fixation) replicate in volumetric medical images. Although the search behaviors associated with global processing ability in volumetric images are not yet well-understood, the measures associated with global processing ability in 2D images serve as a useful starting point for understanding expert search behavior in volumetric tasks. In addition, we investigated how radiologists might establish a global impression of the image using novel measures of scrolling behavior (number of depth passes and scrolling speed). The second aim of this study was to determine how overall search strategy changes with experience. Specially, the goals were to: (1) replicate previous findings that drilling is a better strategy than scanning for lung cancer detection and (2) disentangle the effects of experience from search strategy. Together, these analyses help determine whether existing models of medical image perception can account for expert search behavior in volumetric image interpretation, as well as how they might be updated to account for scrolling behavior in volumetric images.

2 Method

A separate analysis of this dataset has been published previously.[41]

2.1 *Participants*

Fifty-six radiologists were recruited from the National Cancer Institute's Perception Lab at a Radiological Society for North America meeting; a hospital in Salt Lake City, UT, United States; and a hospital in Sydney, NSW, Australia. In order to meet the minimum experience level for eligibility in our study, participants were required to be in the first year of a radiology residency program or higher. Five radiologists were excluded from the study prior to participation due to unsuccessful eye tracking calibration, and data from one radiologist were excluded from the analysis due to equipment failure. The final sample consisted of 50 radiologists with a wide range of experience: 25 radiology residents (4 first year, 5 second year, 7 third year, and 9 fourth year), 1 fellow, and 24 attending or practicing radiologists.

Participants at RSNA were entered into a raffle for a chance to win a $500 Amazon gift card, participants in Salt Lake City were compensated with $50, and participants in Sydney volunteered their time. The study procedures were approved by the University of Utah Institutional Review Board and the Macquarie University Human Research Ethics Committee. All participants provided informed written consent and were debriefed following the study.

2.2 *Procedure*

Participants first completed a questionnaire regarding their level of experience, area of expertise, and demographic information. Next, observers performed a lung cancer detection task using seven axial chest CT scans (one practice and six experimental) viewed in a typical lung window and level. Half of the cases were normal (no lung nodules) and the other half were abnormal (at least one lung nodule). Participants were instructed to identify nodules $\geq$3 mm in diameter by clicking on the nodule's center of mass with the mouse. Case completion time was unrestricted and participants clicked on a box to move on to the next case. Participants could freely scroll back and forth through the slices of the CT scan using the mouse scroll wheel. On average, there were 148 slices in each CT stack. Following each case, radiologists rated the difficulty of the case from 1 (not at all difficult) to 6 (very difficult).

Participants were situated on a chinrest ~89 cm from a 17-arc sec monitor. Eye movements were recorded using an Eyelink 1000 Plus at a sampling rate of 1000 Hz. Participants underwent a nine-point calibration procedure at the beginning of the study, and recalibrations were performed throughout the task as necessary. To reconstruct eye movements through the volumetric space, the observer's current position in depth was co-registered with each eye tracking sample and processed offline using custom MATLAB scripts.

2.3 *Materials*

The abnormal cases contained 9, 11, and 23 nodules, respectively. Five of the six experimental cases were obtained from the Lung Image Database Consortium (LIDC) and the final case was obtained from clinical practice at the University of Utah School of Medicine.[42] For the LIDC cases, ground truth was established by four thoracic radiologists who independently marked nodule locations prior to reviewing the anonymized marks of the other three radiologists and rendering a final decision. For the Utah case, author W.A. marked the nodule locations.

2.4 *Analysis Plan*

The study's sample size, data exclusion criteria, and primary predictions and analyses were preregistered prior to data collection.[43] There are some preregistered analyses that have not yet been conducted as they are beyond the scope of this particular paper (e.g., similarity score and pupillometry). As preregistered, years of experience since graduating medical school and the average number of chest CTs evaluated each week were entered into a multiple linear regression for each of the dependent measures. In addition, in preregistered analyses, image coverage, search strategy (i.e., scanning/drilling), and scrolling speed were regressed onto nodule detection rate to determine which search behaviors predicted better performance. To control for the effects of experience, years of experience and the number of chest CTs read per week were added as predictors in each regression model. The remaining regression analyses were exploratory and not included in the preregistration. We also added a quartile comparison where we compared the bottom and top quartile of each quantitative scanner/driller measure using a between-participants t-test to determine how these methods compared to the subjective method of classifying search strategy.

In addition to the preregistered analyses, Bayes factors were calculated to assist in the interpretation of null results and to help identify analyses that might have been underpowered. For the linear regressions, we used a JZS prior with the default scale ($r = 0.35$). A $BF_{10} > 3$ indicates sufficient evidence for the alternative relative to the null hypothesis, a $BF_{10} < 1/3$ indicates sufficient evidence for the null relative to the alternative hypothesis, and a BF_{10} between these two values indicates that more evidence is needed for a strong conclusion.[44] For each multiple linear regression model, Bayes factors are reported for each predictor variable individually as well as the full model.

3 Results

3.1 *Observer Experience*

Participants (19 females and 31 males) reported reading 41 (SD = 52, range = [0,250]) chest CT scans in an average week and had an average of 12 (SD = 13, range = [0.5, 42]) years of radiology experience since graduating medical school. On average, radiologists were 41 (SD = 13, range = [27,68]) years old. Of these radiologists, 27 (54%) reported they were American Board of Radiology certified or their country's equivalent. Twenty (40%) radiologists reported expertise in thoracic imaging. The relationship between experience and each of the dependent measures of search behavior is shown in Table 1.

Table 1 Results of multiple linear regressions for experience measures.

Measure	Mean	SD	Years p value	Chest CTs p value	β_0	β_1	β_2	Model p value	R^2	BF_{10}
Sensitivity	58%	19%	0.56	0.26	57	−0.001	0.001	0.40	0.04	0.24
False alarms	3.4	2.4	0.21	0.17	3.44	−0.04	0.009	0.14	0.08	0.55
Search time	137.9 s	61.7 s	0.23	0.89	149.1	−0.87	−0.02	0.49	0.03	0.20
Coverage	38%	13%	0.17	0.54	42	−0.002	−0.0002	0.36	0.04	0.26
Saccadic amplitude	2.15 deg	0.77 deg	0.06	0.26	1.85	0.02	0.002	0.12	0.09	0.61
Time to first fixation	567 ms	596 ms	0.24	0.61	501.4	8.25	−0.87	0.40	0.04	0.24
Depth passes	2.3	1.7	0.07	0.22	3.03	−0.04	−0.006	0.11	0.09	0.65
Scrolling speed	6	2	0.004	0.18	7.17	−0.07	−0.007	0.01	0.18	4.59
Refixation rate	39%	11%	0.047	0.27	44	−0.003	−0.0003	0.10	0.10	0.73
Nodule dwell time	3065.8 ms	1531.7 ms	0.60	0.38	3136	9.33	−3.84	0.55	0.03	0.19
Eye movement index	0.41	0.24	0.06	0.32	0.32	0.005	0.001	0.13	0.29	0.60
Change XY/Z score	63.74	64.31	0.02	0.15	31.26	1.79	0.26	0.03	0.14	2.04

3.2 *Task Performance*

3.2.1 *Accuracy*

On average, radiologists reported 58% (SD = 19%) of the lung cancer nodules. Contrary to our prediction, neither years of experience, $F(1,46) = 0.34$, $p = 0.56$, $BF_{10} = 0.36$, nor the number of chest CTs read per week, $F(1,46) = 1.32$, $p = 0.26$, $BF_{10} = 0.54$, predicted nodule detection rate, $R^2 = 0.04$, $BF_{10} = 0.24$ [Fig. 1(a) and Table 1]. Next, we calculated false alarms as the average number of clicks per case that were not within 50 pixels of a true nodule. The average number of false alarms per case was 3.4 (SD = 2.4) nodules. The number of false alarms was not predicted by years of experience, $F(1,46) = 1.62$, $p = 0.21$, $BF_{10} = 0.68$, or the number of chest CTs read per week, $F(1,46) = 1.92$, $p = 0.17$, $BF_{10} = 0.77$; $R^2 = 0.08$, $BF_{10} = 0.55$ [Fig. 1(b) and Table 1]. However, the Bayes factors suggest more evidence is needed to make a conclusion about whether false alarms differ across experience levels.

3.2.2 *Error classification*

Using the eye tracking data, miss errors were classified into recognition, search, or decision errors by calculating the cumulative dwell time on the lung nodules.[45] Recognition errors were defined as unreported nodules fixated for <1000 ms, decision errors were defined as unreported nodules fixated for more than 1000 ms, and search errors were defined as unreported nodules that were never fixated at all. In addition, we performed a non-preregistered analysis to classify search errors into two different types: (1) the slice containing the abnormality was visited but the nodule was never fixated. (2) The slice containing the abnormality was never visited. However, the second type of search error was only observed in 1/50 radiologists, and therefore search errors were collapsed for all subsequent analyses.

Contrary to our prediction, cumulative dwell time on correctly identified nodules ($M = 3065.8$ ms and SD = 1531.7 ms) did not significantly decrease with years of experience, $F(1,46) = 0.27$, $p = 0.60$, $BF_{10} = 0.34$, nor the number of chest CTs read per week, $F(1,46) = 0.79$, $p = 0.38$, $BF_{10} = 0.42$; $R^2 = 0.03$, $BF_{10} = 0.19$. In previous research using a lung cancer detection task with chest radiographs,[45] the distribution of miss errors was 45%

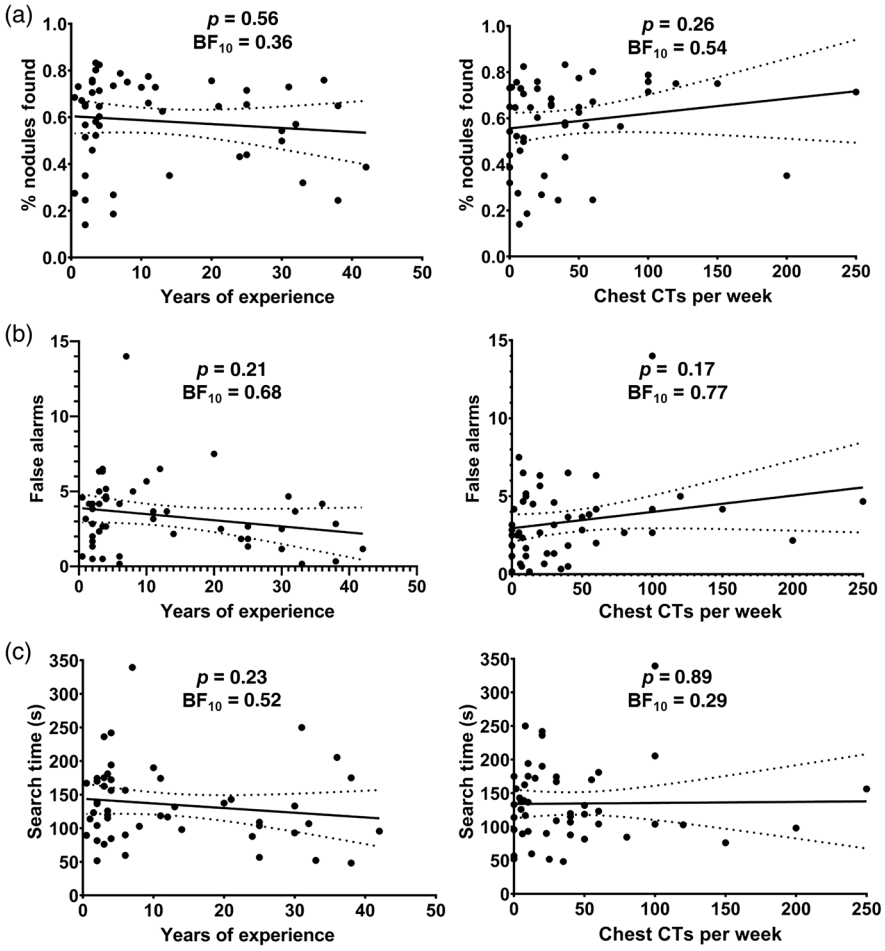

Fig. 1 Relationship between experience and task performance. There was no evidence for a significant relationship between: (a) nodule detection rate and years of experience or the number of chest CTs read per week; (b) false alarms and years of experience or the number of chest CTs read per week; and (c) search time and years of experience or the number of chest CTs read per week. Dashed lines represent the 95% CI here and throughout the manuscript.

decision, 30% search, and 25% recognition errors. In this task, there were 51% recognition errors, 39% search errors, and 10% decision errors. One possible reason for the shift from decision to recognition errors in this dataset is that nodules and normal structures (e.g., blood vessels) might be less confusable with each other in volumetric medical images. However, this proposal will need to be tested in future work by directly comparing 2D and volumetric image search when controlling for other characteristics (e.g., abnormality size and location and case difficulty). Although we predicted the number of search errors would differ with experience, years of experience and the number of chest CTs read per week did not predict a greater proportion of any error type, all p values > 0.05 and all $\mathrm{BF}_{10} < 0.33$.

3.2.3 *Search time*

On average, radiologists spent 137.9 (SD = 61.7) s evaluating each case. Abnormal trials ($M = 163.3$ s, SD = 72.2 s) were searched significantly longer than normal trials ($M = 112.6$ s, SD = 61.5 s), $t(49) = 6.81$, $p < 0.001$, $\mathrm{BF}_{10} = 975782.9$. In 2D images, search time would be expected to decrease with experience due to an increased reliance on the global properties of the image. However, in this volumetric image interpretation task, search time did not decrease with years of experience, $F(1,46) = 1.46$, $p = 0.23$, $\mathrm{BF}_{10} = 0.52$, nor the number of chest CTs read per week, $F(1,46) = 0.02$, $p = 0.89$, $\mathrm{BF}_{10} = 0.29$; $R^2 = 0.03$, $\mathrm{BF}_{10} = 0.20$

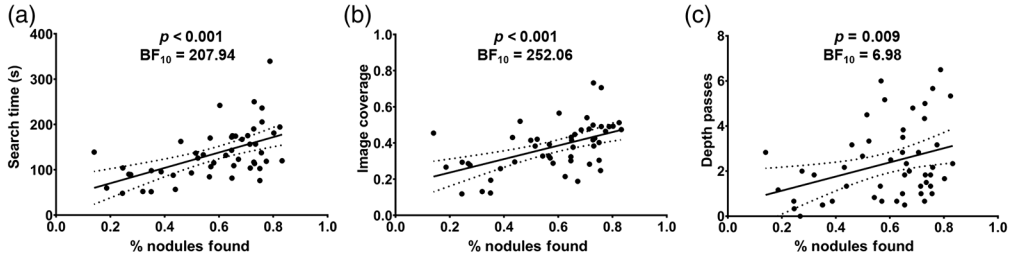

Fig. 2 Predictors of task performance (controlling for experience using multiple linear regression). Higher nodule detection rates were predicted by (a) longer search times, (b) greater image coverage, and (c) more depth passes.

Table 2 Predictors of lung nodule detection rate (controlling for experience using multiple linear regression).

Measure	β	p value	R^2	BF_{10}
Search time	0.002	<0.001	0.31	207.94
Coverage	1.04	<0.001	0.32	252.06
Saccadic amplitude	−0.14	0.02	0.14	3.39
Depth passes	0.05	0.008	0.18	6.98
Scrolling speed	0.04	0.90	<0.001	0.46
Eye movement index	−0.29	0.01	0.17	5.54
Change XY/Z score	−0.0001	0.77	0.04	0.47

[Fig. 1(c) and Table 1]. This pattern of results was the same for both normal and abnormal cases, all p values > 0.05, all $BFs_{10} < 0.33$. Controlling for experience using multiple linear regression, spending more time on each case was a strong predictor of increased nodule detection rate, $F(1,45) = 17.94$, $p < 0.001$, $R^2 = 0.31$, $BF_{10} = 207.94$ [Fig. 2(a) and Table 2].

3.3 *Eye Movements*

3.3.1 *Image coverage*

To calculate the percentage of lung tissue searched (i.e., image coverage), each slice of the CT scans was converted to a black (non-lung tissue) and white (lung tissue) mask. Using the eye tracking sample data, which consisted of the x, y, and z eye position coordinates sampled once every millisecond, we converted the pixels within a 2.6-deg diameter UFOV of each set of coordinates to black. None of the results reported here substantively differ if coverage is calculated using the fixation data instead of the eyetracking sample data. Although a 5-deg diameter UFOV is commonly used for lung nodule detection tasks using chest radiographs,[46] previous research demonstrated a 2.6-deg diameter UFOV is more appropriate for lung nodule detection using chest CT scans.[47] Image coverage was calculated as [1 − (the number of white pixels in the final image/the number of white pixels in the original image)].

Consistent with previous research using volumetric medical images, overall image coverage was quite low.[36,38,48,49] On average, only 38% (SD = 13%) of the total area of the CT scans was searched within a 2.6-deg diameter UFOV. We predicted that image coverage would decrease with observer experience, indicating an ability to rely more on information extracted from the periphery rather than a systematic search. Contrary to this prediction, image coverage did not decrease with years of experience, $F(1,46) = 1.91$, $p = 0.17$, $BF_{10} = 0.58$, nor the number of

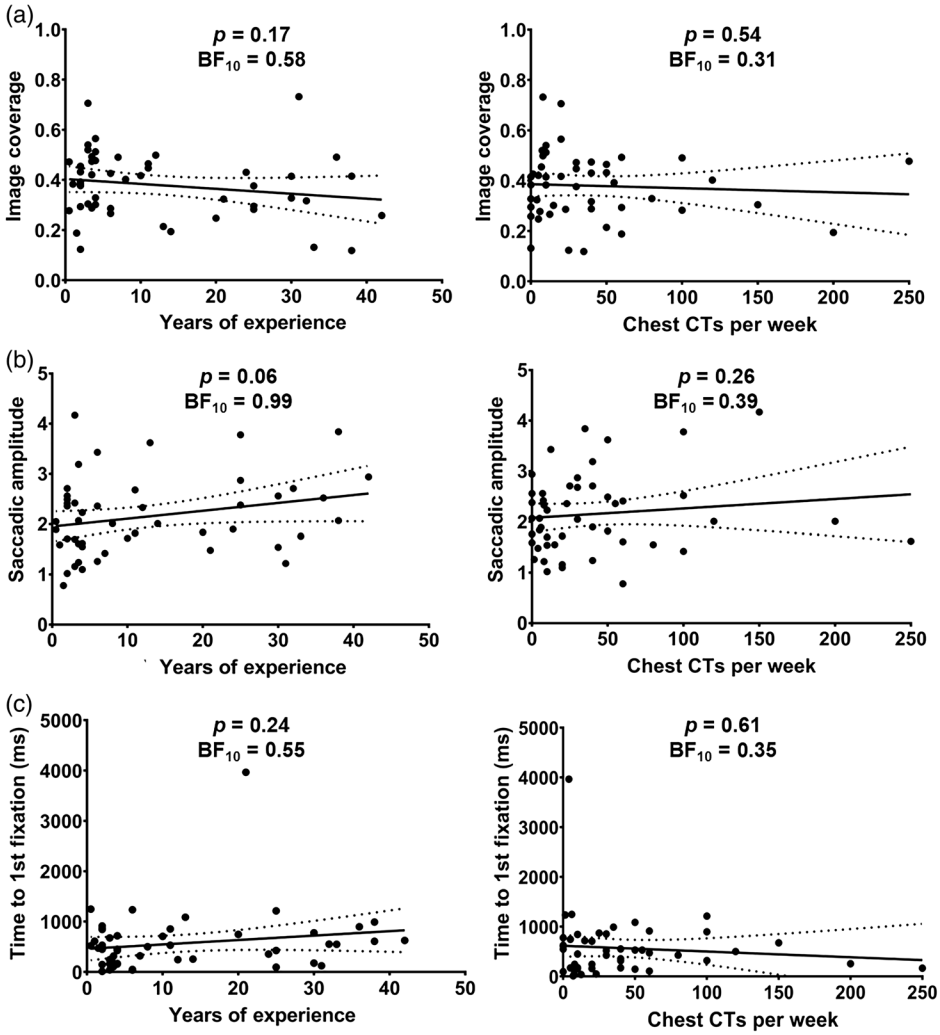

Fig. 3 Relationship between experience and search behavior. There was no evidence for a significant relationship between (a) image coverage and years of experience or the number of chest CTs read per week; (b) saccadic amplitude and years of experience or the number of chest CTs read per week; and (c) time to first fixation and years of experience or the number of chest CTs read per week.

chest CTs evaluated each week, $F(1,46) = 0.39$, $p = 0.54$, $\mathrm{BF}_{10} = 0.31$; $R^2 = 0.04$, $\mathrm{BF}_{10} = 0.26$ [Fig. 3(a) and Table 1]. Controlling for experience using multiple linear regression, searching the images more thoroughly strongly predicted increased nodule detection rate, $F(1,45) = 18.57$, $p < 0.001$, $R^2 = 0.32$, $\mathrm{BF}_{10} = 252.06$ [Fig. 2(b) and Table 2].

3.3.2 *Saccadic amplitude*

A larger saccadic amplitude (i.e., the average distance between consecutive fixations expressed in degrees of visual angle) is thought to reflect a more global search strategy and was expected to increase with observer experience.[10] On average, saccadic amplitude was 2.15 deg (SD = 0.77 deg). Contrary to our prediction, saccadic amplitude did not significantly increase with years of experience, $F(1,46) = 3.61$, $p = 0.06$, $\mathrm{BF}_{10} = 0.99$, nor the average number of chest CTs read per week, $F(1,46) = 1.29$, $p = 0.26$, $\mathrm{BF}_{10} = 0.39$; $R^2 = 0.09$, $\mathrm{BF}_{10} = 0.61$ [Fig. 3(b) and Table 1]. However, the Bayes factors suggest more evidence is needed before a strong conclusion can be made about the relationship between experience and saccadic amplitude. Controlling for experience using multiple linear regression, having a smaller saccadic

amplitude predicted a higher nodule detection rate, $F(1,45) = 5.45$, $p = 0.02$, $R^2 = 0.14$, $BF_{10} = 3.39$ (Table 2).

3.3.3 *Time to first fixation*

Time to first fixation on a detected abnormality in 2D medical images is thought to reflect a more global search strategy and typically decreases with experience.[10] To adapt this measure to volumetric images, we calculated time to first fixation relative to the moment the abnormality first became visible when scrolling through the slices.[50] If the nodule was not detected the first time, it became visible (i.e., the radiologist moved to another position in depth without clicking on the nodule), time to first fixation was calculated relative to the moment the abnormality first reappeared prior to detection.

On average, radiologists took 567 (SD = 596) milliseconds to fixate on the nodules from the moment they first became visible. Contrary to our prediction, time to first fixation did not decrease with years of experience, $F(1,46) = 1.41$, $p = 0.24$, $BF_{10} = 0.55$, nor the number of chest CTs read per week, $F(1,46) = 0.27$, $p = 0.61$, $BF_{10} = 0.35$; $R^2 = 0.04$, $BF_{10} = 0.24$ [Fig. 3(c) and Table 1]. Upon visual inspection [Fig. 3(c)], it became apparent that one participant was an outlier (>3 SDs from the mean). However, the outcome of the multiple linear regression does not change if this outlier is removed, both p values > 0.05, all $BFs_{10} < 0.55$.5

3.3.4 *Refixation rate*

Refixation rate was calculated as the proportion of total fixations that were within UFOV (2.6 deg) of a previous fixation (i.e., proportion of fixations that were refixations).[51] We predicted that more experienced radiologists would use more systematic search strategies to navigate through the image, resulting in fewer refixations.

On average, 39% (SD = 11%) of fixations were refixations. In partial support for our hypothesis, refixation rate decreased with years of experience, $F(1,46) = 4.16$, $p = 0.047$, $BF_{10} = 1.21$, but not the number of chest CTs read per week, $F(1,46) = 1.24$, $p = 0.27$, $BF_{10} = 0.37$; $R^2 = 0.10$, $BF_{10} = 0.73$. However, the Bayes factors indicate that more evidence is needed to make a strong conclusion about the relationship between refixations and observer experience. Controlling for experience using multiple linear regression, higher refixation rates predicted better nodule detection performance, $F(1,45) = 5.67$, $p = 0.02$; $R^2 = 0.15$, $BF_{10} = 3.67$. This result suggests that observers with larger refixation rates may benefit from additional opportunities to detect nodules that might have been missed during the first opportunity for detection. Consistent with this proposal, refixation rate was strongly correlated with search time, $F(1,48) = 62$, $p < 0.001$, $r^2 = 0.56$, $BF_{10} = 3.57 \times 10^7$, and higher refixation rates were no longer significantly associated with better performance when controlling for search time, $F(1,47) = 1.04$, $p = 0.31$; $R^2 = 0.30$, $BF_{10} = 0.41$.

3.4 *Scrolling Behavior*

3.4.1 *Depth passes*

The number of passes through the depth of the CT scan has been proposed as a metric of global processing ability in volumetric images.[31] If experienced observers rely more on a global search strategy, they may make more passes through the depth of the image in order to establish a global impression of the image. Alternatively, if more experienced observers are able to extract the global properties of the image more easily, they might be able to maintain high-performance despite making fewer passes through the depth of the image. The number of passes through depth was defined as the number of times the radiologist scrolled through at least 80% of the depth of the full stack.

On average, radiologists made 2 (SD = 2) depth passes. Contrary to our prediction, the number of passes through depth was not significantly related to years of experience, $F(1,46) = 3.57$, $p = 0.07$, $BF_{10} = 0.95$, nor the number of chest CTs read per week, $F(1,46) = 1.53$, $p = 0.22$,

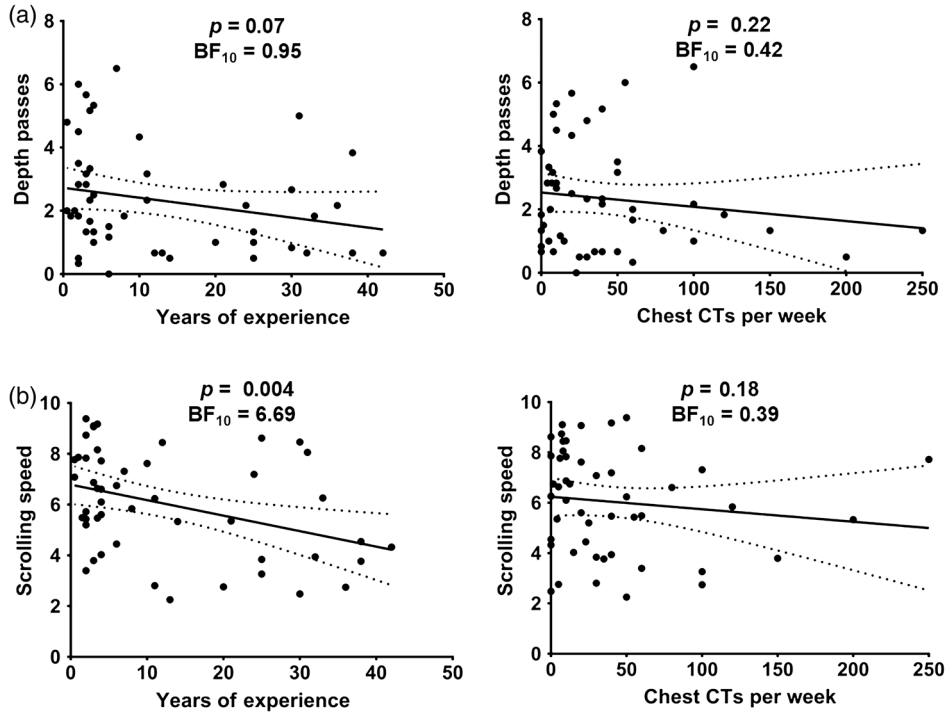

Fig. 4 Relationship between experience and scrolling behavior. (a) There was no significant relationship between the mean number of depth passes per case and years of experience or the number of chest CTs read per week. Depth passes are defined as the number of times the radiologist scrolled through at least 80% of the CT scan. (b) Scrolling speed (slices per second) decreased with years of experience but did not differ with the number of chest CTs read per week.

$BF_{10} = 0.42$; $R^2 = 0.09$, $BF_{10} = 0.65$ [Fig. 4(a) and Table 1]. However, Bayes factors suggest that more evidence is needed before making a strong conclusion about the relationship between the number of passes through depth and experience. Controlling for experience using multiple linear regression, making more passes through depth predicted increased nodule detection rate, $F(1,45) = 7.51$, $p = 0.009$; $R^2 = 0.18$, $BF_{10} = 6.98$ [Fig. 2(c) and Table 2].

3.4.2 *Scrolling speed*

On average, scrolling speed was 6 (SD = 2) slices per second. We predicted that more experienced observers would scroll through the stack more quickly than less experienced radiologists. However, contrary to this prediction, scrolling speed significantly decreased with years of experience, $F(1,46) = 9.14$, $p = 0.004$, $BF_{10} = 6.69$ but not the number of chest CTs read per week, $F(1,46) = 1.82$, $p = 0.18$, $BF_{10} = 0.39$; $R^2 = 0.18$, $BF_{10} = 4.59$ [Fig. 4(b) and Table 1]. Controlling for experience using multiple linear regression, scrolling speed did not predict differences in nodule detection rate, $F(1,45) = 0.02$, $p = 0.90$, $R^2 < 0.001$, $BF_{10} = 0.46$ (Table 2).

3.5 *Scanners and Drillers*

Radiologists were first tentatively divided into scanners and drillers by analyzing the depth by time plots for each participant following the subjective method used in the previous studies [Fig. 5(a)].[36,39] First, the observer's position in depth was plotted on the y axis and time was plotted on the x axis. Next, each quadrant of the image was assigned a different color. At each time point, the observer's eye position on the 2D plane was reduced to a single dimension by plotting each point in the color assigned to that quadrant. Using the depth by time plots, the lead author then made a subjective decision about whether each radiologist was a driller or a scanner according to the descriptions of search strategy outlined by Drew et al. (2013). Qualitatively,

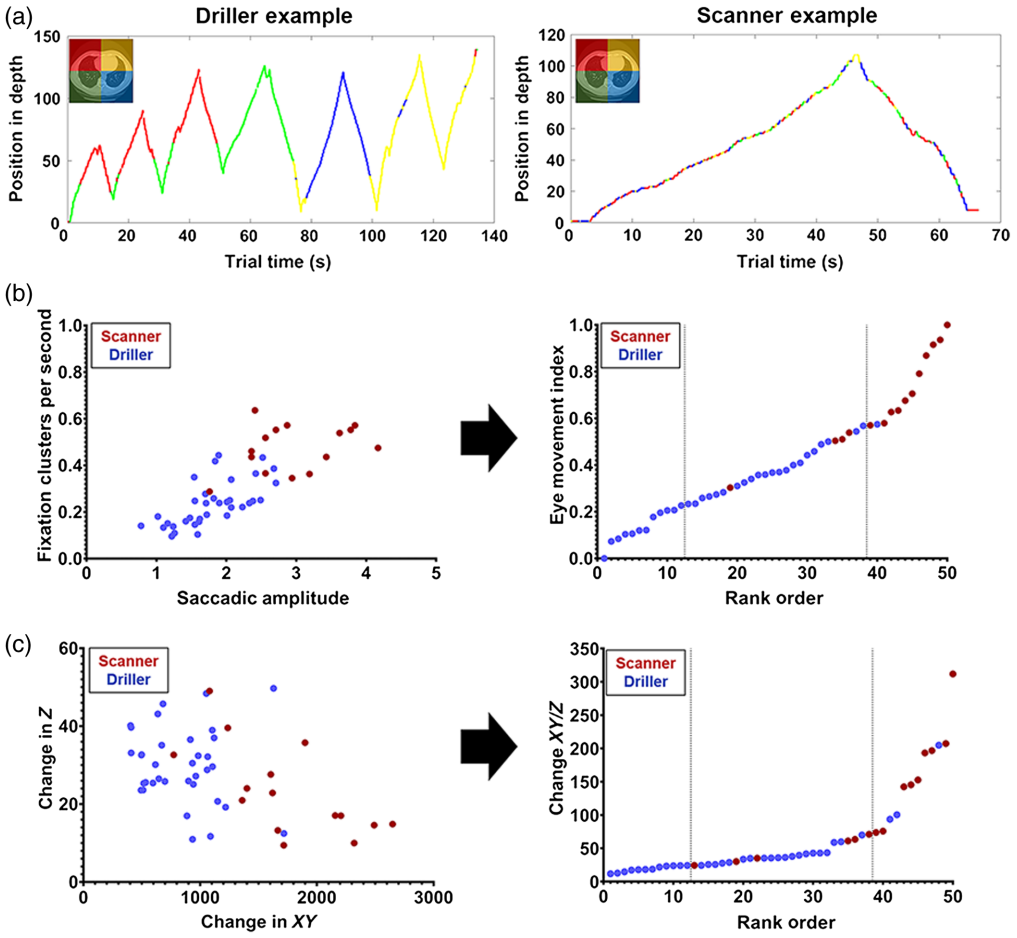

Fig. 5 Different methods of categorizing scanners and drillers: (a) depth by time plots; (b) eye movement index = normalized saccadic amplitude + normalized number of fixation clusters per second; and (c) change in XY/Z = summed scan path distance/change in Z averaged over 5-s intervals. The color coding in (b) and (c) reflects the groups determined using the subjective categorization method described in (a). Dashed lines represent the dividing point for the lower and upper quartiles.

driller plots are characterized by spending prolonged time in one region of the lung (typically one quadrant or lobe) at a time while rapidly scrolling through the slices. In contrast, scanners search broadly across the 2D plane while slowly moving through the depth of the CT scan [Fig. 5(a)]. Although depth by time plots can reveal qualitative differences in search strategy, it is unclear how to best capture these differences in search behavior quantitatively. Here we compared two quantitative measures that have been used in the previous research: the eye-movement index[36,39] and the change in XY/Z score.[38]

In the original scanner/driller study, the authors' subjective categorizations of search strategy were then tested using the eye movement index.[36] On average, scanners should have larger saccadic amplitude and make fewer consecutive fixations in the same quadrant of the lung (i.e., fixation clusters) than drillers. Therefore, if mean saccadic amplitude is plotted on the x axis and the average number of fixation clusters per second is plotted on the y axis, scanners tend to cluster in the top-right of the figure [Fig. 5(b)]. These measures can then be combined into a single metric by normalizing each score from 0 to 1 and adding the two measures together [Fig. 5(b)].

The eye movement index can help distinguish between scanners and drillers,[36] but this metric does not directly take the observer's movement through depth into account. If drilling is associated with better performance because it enables radiologists to take advantage of abrupt motion onset cues while scrolling through depth, this may be an important aspect of search behavior to

quantify. To account for this possibility, scanning and drilling behavior has also been conceptualized as the [summed change in xy (i.e., scan path length)/the maximum change in z] averaged across 5-s intervals.[38] Within a set time period, drillers make more movements in Z than in XY compared to scanners, resulting in smaller change in XY/Z scores than scanners [Fig. 5(c)]. Another promising approach is to classify scanners and drillers based on the number of direction changes that occur during each case.[52] However, this measure requires a fixed time limit for each CT scan, so we were not able to use this categorization method for the current dataset.

Both EMI and change in XY/Z scores have been used in the previous research, but there is no consensus on which best captures the qualitative differences in search strategy observed in depth by time plots. Although there is some overlap in these measures, an observer can still score relatively high on one and relatively low on the other, suggesting they tap into distinct aspects of search behavior [Fig. 5).[36,38] Furthermore, it is unclear if search strategy is dichotomous (e.g., scanners versus drillers), or whether it is more appropriate to consider continuous changes in these measures (e.g., more drilling versus less drilling behavior). Here we used the eye movement index and change in XY/Z scores as continuous predictors for each of the dependent variables using linear regression. In addition, we also divided radiologists into groups based on quartile rankings and compared these results to the subjective categorization method described above. The subjective categorization method and the change in XY/Z score regression analyses were preregistered,[43] but the eye movement index and quartiles analyses were exploratory.

3.5.1 *Subjective categorization method*

Using the subjective categorization method [Fig. 5(a)], 30% of radiologists were categorized as scanners and 70% were categorized as drillers. We first present the results using this separation and then examine the degree to which the different objective methods of quantifying search strategy impact the results.

Controlling for experience using multiple linear regression, drillers ($M = 65\%$ and $\mathrm{SD} = 14\%$) detected more of the lung nodules than scanners ($M = 44\%$ and $\mathrm{SD} = 20\%$), $F(1,45) = 16.41$, $p < 0.001$, $\mathrm{BF}_{10} = 129.67$ [Fig. 6(a)]. Drillers ($M = 4.1$, $\mathrm{SD} = 2.4$) also made more false alarms per case than scanners ($M = 1.7$, $\mathrm{SD} = 1.4$), $F(1,45) = 12.28$, $p = 0.001$, $\mathrm{BF}_{10} = 35.50$, but it is possible that some true nodules may be unmarked in the LIDC database (see also Ref. 15) so we do not want to over-emphasize false alarms. Scanners ($M = 48\%$, $\mathrm{SD} = 20\%$) made significantly more search errors than drillers ($M = 35\%$, $\mathrm{SD} = 15\%$), $t(48) = 2.67$, $p = 0.01$, $\mathrm{BF}_{10} = 4.67$, whereas drillers ($M = 54\%$, $\mathrm{SD} = 12\%$) made significantly more recognition errors than scanners ($M = 43\%$, $\mathrm{SD} = 17\%$), $t(48) = 2.73$, $p = 0.009$, $\mathrm{BF}_{10} = 5.28$. There were no significant differences between scanners ($M = 8\%$, $\mathrm{SD} = 7\%$) and drillers ($M = 11\%$, $\mathrm{SD} = 8\%$) on decision errors, $t(48) = 0.83$, $p = 0.41$, $\mathrm{BF}_{10} = 0.37$.

These large differences in hit rate between the search strategies were not associated with differences in years of experience, $t(48) = 1.16$, $p = 0.25$, $\mathrm{BF}_{10} = 0.49$, nor the number of chest CTs read per week, $t(48) = 0.14$, $p = 0.89$, $\mathrm{BF}_{10} = 0.29$ [Fig. 6(a)]. What did seem to drive the improved hit rate was that drillers spent more time evaluating each case, $t(48) = 3.23$, $p = 0.002$, $\mathrm{BF}_{10} = 15.71$, searched the images more thoroughly, $t(48) = 4.29$, $p < 0.001$, $\mathrm{BF}_{10} = 252.04$, and made more passes through depth, $t(48) = 2.23$, $p = 0.03$, $\mathrm{BF}_{10} = 2.06$, than scanners.

Using the subjective categorization, we then examined the eye movement index and change in XY/Z scores for the two groups. On average, scanners ($M = 0.68$, $\mathrm{SD} = 0.19$) had a larger eye movement index than drillers ($M = 0.30$, $\mathrm{SD} = 0.15$), $t(48) = 7.55$, $p < 0.001$, $\mathrm{BF}_{10} = 6.64 \times 10^6$; and scanners ($M = 118.9$, $\mathrm{SD} = 82.6$) had a larger change in XY/Z score than drillers ($M = 40.1$, $\mathrm{SD} = 35.1$), $t(48) = 4.77$, $p < 0.001$, $\mathrm{BF}_{10} = 956.03$.

We then examined the effect of using quantitative categorizations, repeating the above analyses using EMI and XY/Z quantitative measures as: (1) continuous predictors of performance and (2) to classify radiologists into distinct groups of scanners and drillers using the top and bottom quartiles, respectively.

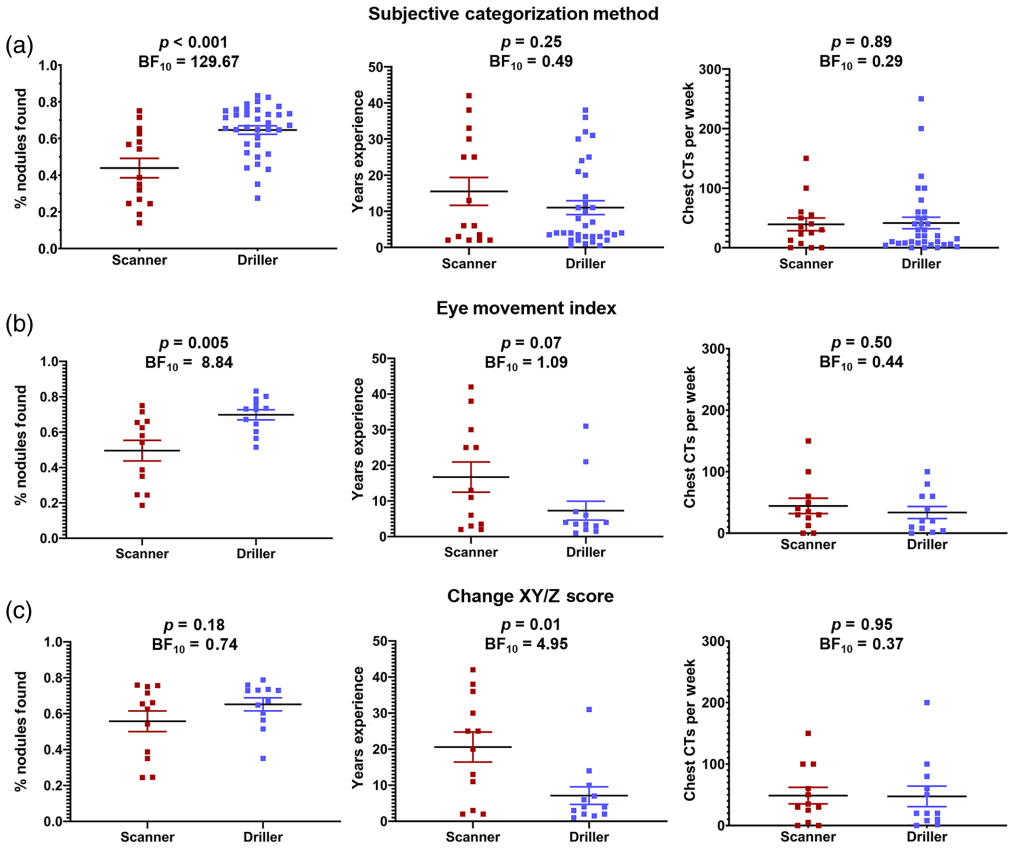

Fig. 6 The relationship between search strategy, experience, and task performance using different methods of classifying search strategy. (a) Subjective categorization using depth by time plots; (b) categorization using the eye movement index by quartiles; (c) categorization using the change in XY/Z scores by quartiles. The solid line indicates the mean value, dots represent the individual data points, and error bars represent standard error of the mean.

3.5.2 *Eye movement index*

First, we used the eye movement index [Fig. 5(b)] as a continuous predictor of performance in a linear regression analysis. Controlling for experience using multiple linear regression, having a smaller eye movement index (drilling) was associated with better nodule-detection rates than having a large eye movement index (scanning), $F(1,45) = 6.85$, $p = 0.01$, $R^2 = 0.17$, $\mathrm{BF}_{10} = 5.54$ [Figs. 7(a) and Table 2]. Next, we sought to determine whether these measures could be used to establish an objective classification system by dividing radiologists into scanners and drillers using the top and bottom quartiles, respectively. Using this method, 12/12 radiologists in the bottom quartile matched our subjective "drilling" classification, and 11/12 radiologists in the top quartile matched our "scanning" classification [Fig. 5(b)]. If we then look at the performance of these two quartile groups on the nodule detection, the drillers (bottom quartile) detected 70% (SD = 10%) of the nodules, on average, whereas the scanners (top quartile) detected only 50% (SD = 20%) of the nodules, $t(22) = 3.13$, $p = 0.005$, $\mathrm{BF}_{10} = 8.84$ [Fig. 6(b)]. The distribution of error type follows the same pattern as the subjectively categorized results: scanners ($M = 47\%$, SD = 20%) made significantly more search errors than drillers ($M = 30\%$, SD = 12%), $t(22) = 2.56$, $p = 0.02$, $\mathrm{BF}_{10} = 3.37$, whereas drillers ($M = 57\%$, SD = 8%) made significantly more recognition errors than scanners ($M = 44\%$, SD = 18%), $t(22) = 2.26$, $p = 0.03$, $\mathrm{BF}_{10} = 2.16$. There were no significant differences between scanners ($M = 14\%$, SD = 7%) and drillers ($M = 9\%$, SD = 8%) on decision errors, $t(22) = 1.59$, $p = 0.13$, $\mathrm{BF}_{10} = 0.92$.

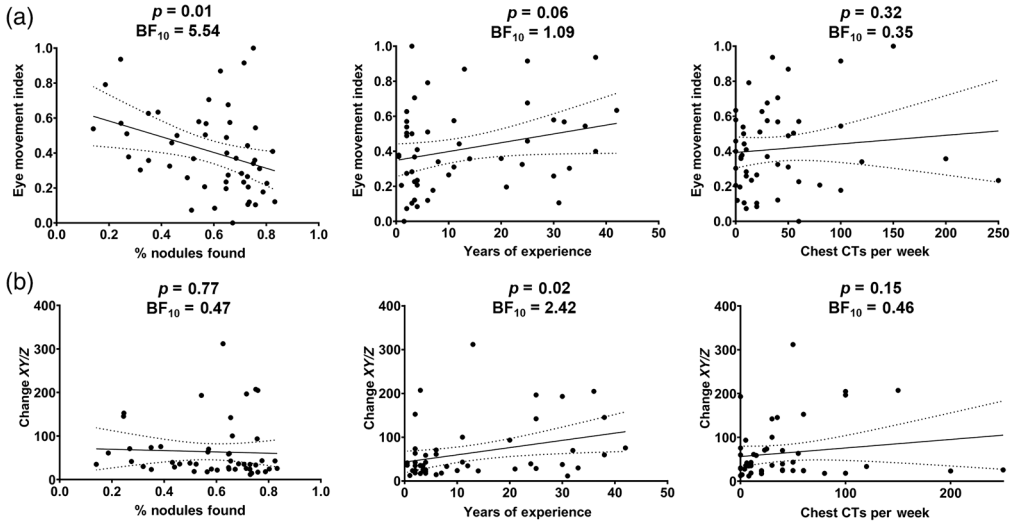

Fig. 7 The relationship between search strategy, experience, and task performance using different methods of classifying search strategy. (a) Eye movement index: lower scores reflect more drilling behavior. (b) Change in XY/Z scores: lower scores indicate more drilling behavior.

Neither years of experience, $F(1,46) = 3.79$, $p = 0.06$, $\mathrm{BF}_{10} = 1.09$ (although note the insufficient evidence here), nor the number of chest CTs per week, $F(1,46) = 1.02$, $p = 0.32$, $\mathrm{BF}_{10} = 0.35$, predicted the eye movement index, $R^2 = 0.29$, $\mathrm{BF}_{10} = 0.60$ [Fig. 7(a) and Table 1]. The bottom ($M = 7$, SD = 9) and top ($M = 17$, SD = 15) quartiles did not significantly differ in years of experience, $t(22) = 1.89$, $p = 0.07$, $\mathrm{BF}_{10} = 1.30$ [Fig. 6(b)], but the Bayes factor indicates that more evidence is needed to make a strong conclusion. Similarly, the bottom ($M = 34$, SD = 34) and top ($M = 44$, SD = 43) quartiles did not differ in the number of chest CTs read per week, $t(22) = 0.68$, $p = 0.50$, $\mathrm{BF}_{10} = 0.44$ [Fig. 6(b)].

Using the eye movement index as a continuous measure, we found that drilling was associated with longer search times, $F(1,48) = 11.41$, $p = 0.002$, $\mathrm{BF}_{10} = 22.79$, greater image coverage, $F(1,48) = 12.94$, $p = 0.001$, $\mathrm{BF}_{10} = 39.29$, and more depth passes, $F(1,48) = 7.42$, $p = 0.009$, $\mathrm{BF}_{10} = 5.27$. As seen in the subjective classification method, the quartile analysis revealed that drillers (bottom quartile) spent more time evaluating each case, $t(22) = 3.2$, $p = 0.004$, $\mathrm{BF}_{10} = 9.93$, searched the images more thoroughly, $t(22) = 3.47$, $p = 0.002$, $\mathrm{BF}_{10} = 16.48$, and made more passes through depth, $t(22) = 2.90$, $p = 0.008$, $\mathrm{BF}_{10} = 5.96$, than scanners [top quartile, Fig. 8(a)].

3.5.3 *Change in XY/Z score*

Next, we used the change in XY/Z scores as our key variable [Fig. 5(c)].[38] Controlling for experience using multiple linear regression, change in XY/Z scores did not significantly predict nodule detection rate, $F(1,45) = 0.09$, $p = 0.77$, $R^2 = 0.04$, $\mathrm{BF}_{10} = 0.47$ [Fig. 7(b) and Table 2]. Using the quartile method, 12/12 radiologists in the bottom quartile matched our subjective "drilling" classification, and 9/12 radiologists in the top quartile matched our "scanning" classification [Fig. 5(c)]. Drillers detected 65% (SD = 13%) of the nodules, whereas scanners detected 56% (SD = 20%) of the nodules. These differences were not statistically significant, $t(22) = 1.39$, $p = 0.18$, $\mathrm{BF}_{10} = 0.74$ [Fig. 6(c)], but the Bayes factors indicate there is insufficient evidence to interpret these null findings. For the change in XY/Z score, there were no significant differences in the type of miss errors between scanners and drillers, all p values > 0.05.

Radiologists with a larger change in $XY/ZXY/Z$ score (scanners) tended to have more years of experience, $F(1,46) = 6.4$, $p = 0.02$, $\mathrm{BF}_{10} = 2.42$, but there was no relationship between change in XY/Z score and the number of chest CTs read per week, $F(1,46) = 2.20$,

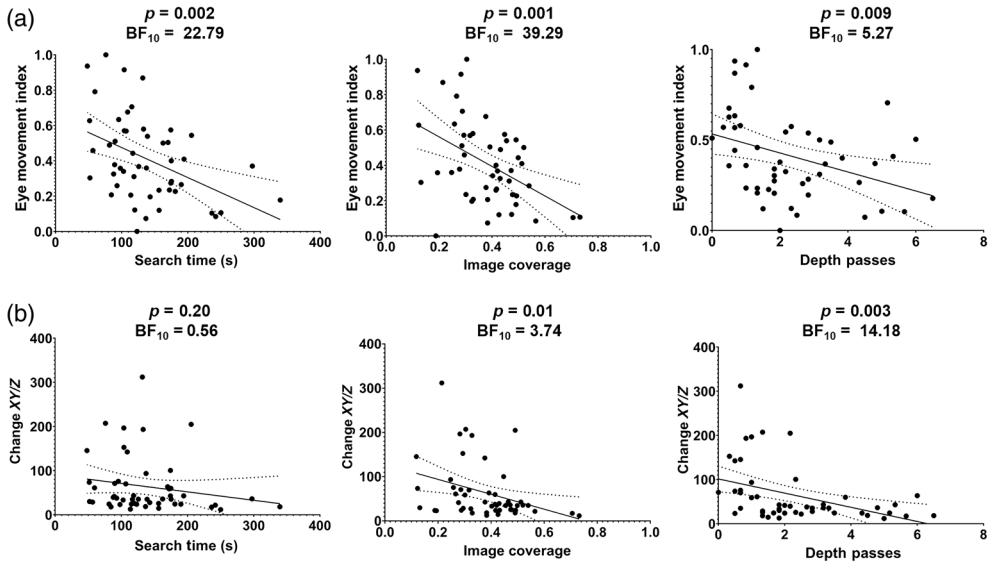

Fig. 8 Scanner and driller search behavior. (a) Eye movement index: lower scores reflect greater drilling behavior. (b) Change in XY/Z score: lower scores reflect greater drilling behavior.

$p = 0.15$, $\mathrm{BF}_{10} = 0.46$; $R^2 = 0.14$, $\mathrm{BF}_{10} = 2.04$ [Fig. 7(b) and Table 1]. Drillers ($M = 7$, $\mathrm{SD} = 9$) had fewer years of experience than scanners ($M = 21$, $\mathrm{SD} = 14$), $t(22) = 2.79$, $p = 0.01$, $\mathrm{BF}_{10} = 4.95$, but the bottom ($M = 48$, $\mathrm{SD} = 58$) and top ($M = 49$, $\mathrm{SD} = 47$) quartiles did not differ in the number of chest CTs read per week, $t(22) = 0.06$, $p = 0.95$, $\mathrm{BF}_{10} = 0.37$ [Fig. 6(c)].

Using change in XY/Z scores as a continuous measure, drilling was associated with greater image coverage, $F(1,48) = 6.52$, $p = 0.01$, $\mathrm{BF}_{10} = 3.74$, and more depth passes, $F(1,48) = 10.10$, $p = 0.003$, $\mathrm{BF}_{10} = 14.18$ but was not significantly related to search time, $F(1,48) = 1.67$, $p = 0.20$, $\mathrm{BF}_{10} = 0.56$. Similarly, in the quartile analysis, drillers had longer search times, $t(22) = 2.11$, $p = 0.046$, $\mathrm{BF}_{10} = 1.74$, greater image coverage, $t(22) = 2.49$, $p = 0.02$, $\mathrm{BF}_{10} = 3.05$, and more depth passes, $t(22) = 3.87$, $p < 0.001$, $\mathrm{BF}_{10} = 35.24$, than scanners [Fig. 8(b)].

3.5.4 *Results summary*

Across the three methods of classifying search behavior, our results largely replicate previous findings that drilling is a superior strategy for lung nodule detection than scanning when controlling for the effects of experience. Both the subjective categorization method and the eye movement index revealed greater nodule detection for drilling than scanning. The change in XY/Z score did not significantly predict performance; however, the Bayes factors indicate that these analyses are not interpretable with this sample size. Critically, this study expands on the previous research by examining whether differences in experience level between the two groups can account for differences in performance. On average, drillers tended to have less experience than scanners [Fig. 7(b)], which is inconsistent with the idea that radiologists learn to adopt better search strategies with experience. However, this data should not be interpreted as evidence that more experienced observers are worse at the task overall. We do not see any evidence for a negative relationship between experience and detection rate in our dataset [Fig. 1(a)], and there are many additional factors that may explain variation in task performance beyond search strategy. Rather, these results demonstrate that drilling behavior predicts better performance above and beyond the effects of experience. Drillers may have performed better on the task because they engaged in a more systematic search of the images: regardless of how we classified the radiologists, drilling was associated with greater image coverage, making more passes through depth, and spending more time on each case.

4 Discussion

In this study, we examined how naturalistic search behavior differed across radiologists with varying levels of experience during lung cancer detection with volumetric images. This research makes two primary contributions to the literature. First, contrary to predictions based on findings from studies using 2D medical images, we did not find evidence in support of global processing-related changes in search behavior with experience—and, importantly, we demonstrate evidence for the null using Bayes analyses. Null results were consistent across a number of measures that have been closely associated with expertise in 2D medical image interpretation (search time, image coverage, saccadic amplitude, and time to first fixation) as well as novel measures of scrolling behavior (depth passes and scrolling speed) that have been proposed as potential indices of expertise in volumetric image interpretation.[31,33] Second, we identified several strong predictors of individual differences in task performance for lung cancer detection. Although experts tend to have better performance than novices in 2D interpretation tasks despite lower image coverage, we found that performance in our volumetric task was closely related to how many opportunities there were for abnormality detection. Specifically, better performance was predicted by spending more time on each case, searching the images more thoroughly, and making more passes through the depth of the CT scan. Observers who adopted a drilling search strategy detected more of the lung cancer nodules than scanners, which may be due to differences in how systematically the images were searched. Critically, these performance differences do not appear to be driven by differences in experience level. Drilling remained a significant predictor of the performance when controlling for differences in experience, and there was limited evidence that drillers actually had fewer years of experience than scanners. Together, these findings have important implications for current models of perceptual expertise and may provide insight on how to train radiologists to evaluate volumetric images.

Although this research suggests a smaller role for global processing in volumetric image interpretation than in 2D images, these results need to be reconciled with recent reports that radiologists can reliably classify volumetric images as normal or abnormal after brief video presentations.[29,30] The current study used lung cancer detection rather than the breast cancer and prostate cancer detection tasks used in the previous studies, suggesting differences in stimulus characteristics (e.g., abnormality size) might account for the different findings. In addition, the type of signal that supports abnormality detection in gist processing studies could be quite different in volumetric images, where a global "snapshot" of the image is not present. Instead, the abrupt motion onset cues elicited by abnormalities in the periphery as the videos transition through depth might be the key driver for performance, rather than sensitivity to global scene statistics *per se*. In future research, it may be fruitful to determine how different abnormality characteristics, such as their ability to elicit motion onset cues, relate to performance in gist processing studies. In previous flash-viewing studies, radiologists were able to detect cancerous "signals" in the breast opposite to the lesion, as well as images taken years before the development of a detectable mass.[53,54] If the presence of a mass is also unnecessary for gist processing in volumetric images, it would suggest that the outcome of previous gist processing studies did not depend solely on motion onset cues that may have been generated by the abnormalities when the videos transitioned through depth.

A clearly plausible explanation for global processing playing a smaller role in volumetric rather than 2D image interpretation is that the global statistical properties of the image cannot be extracted in a single glance and must instead be acquired as the observer scrolls through the depth of the stack. If the gist of the image is not readily available, it might then become more important to rely on a more systematic, foveal search through the image, which is a characteristic of drilling behavior. Consistent with this interpretation, many of our current results show the opposite relationship between scan patterns and task performance than would be expected under global processing models. Specifically, nodule detection rate was strongly predicted by how thoroughly the images were searched, suggesting less information can be extracted from the periphery during volumetric image interpretation. Notably, this is consistent with the recent work demonstrating that UFOV is lower for lung cancer detection in volumetric medical images than UFOV estimates established for the same task using chest radiographs.[47]

Searching medical images more thoroughly might be particularly important in volumetric images due to their large size (when taking depth into account). Because volumetric images consist of hundreds of stacked 2D images,[26,28] abnormalities represent a smaller fraction of the total image size and are only visible for a brief period of the overall search time.[47] In addition, image coverage tends to be much lower in volumetric medical images than in 2D medical images, and recent work suggests using volumetric images may not be beneficial if abnormalities cannot be readily detected using peripheral vision.[38,48,49] Image size is a particularly important consideration in light of findings from the visual search literature that demonstrate that memory for where you have already searched is very limited. At best, observers are able to remember their most recent 3 to 4 fixations or have a rough representation of their general scan path.[55–58] At worst, observers are no better than chance at distinguishing their own scan patterns from someone else's following a visual search task.[59–61] Finally, memory appears to be impaired following brief interruptions in volumetric image interpretation tasks,[51,62] suggesting memory representations might be easily disrupted in stacked images. Thus previous findings suggest that it might be particularly difficult to maintain a reliable representation of which regions of the image have already been searched in volumetric images, and the current work suggests that optimally deciding when to terminate search may be a strong predictor of individual differences in the performance.

Although this discussion highlights the potential costs of searching through stacked images, this should not be considered criticism of using volumetric images in radiology. Volumetric medical images are associated with better overall diagnostic accuracy across a wide range of clinical tasks and allow radiologists to more easily envision the underlying 3D nature of anatomical structures and abnormalities.[37,38,48,63,64] The current results demonstrate there may be an optimal strategy for evaluating volumetric medical images in clinical practice. In the previous research, rapidly drilling through the slices appeared to be a better strategy for lung nodule detection than scanning the 2D plane while slowly moving through depth in a time limited study (3 min per case).[36] Here we replicated these findings while allowing observers an unlimited amount of time to evaluate each case, demonstrating that it is not simply that scanning is a less efficient method, but that it actually does lead to more miss errors. Most importantly, our findings demonstrate these effects were not driven by differences in level of experience: scanners and drillers had a similar number of years of experience and chest CTs evaluated per week. Moreover, drilling behavior remained a significant predictor of task performance when controlling for observer experience. Thus it may be beneficial for radiologists to adopt a drilling strategy when evaluating chest CT images for lung nodules.

Although the benefits of drilling may be due to the ability to elicit abrupt motion onset cues when rapidly scrolling through depth,[37] there are other potential explanations for the observed differences in the performance. For example, using this dataset alone, we cannot rule out that the drillers in our study might have been more conscientious or motivated than scanners, on average, resulting in both better performance and a more thorough search of the images. However, the previous research found that teaching radiology residents to use a drilling strategy improved task performance, which suggests the benefits of drilling cannot solely be a result of group-level differences between observers.[65] Alternatively, drilling might be a more effective strategy because it reflects a more systematic approach to searching through volumetric images. Drillers tend to search through one lobe or quadrant of the lung before moving on to a next one, which appears to result in a more thorough search of the image. In contrast, scanners' search patterns appear to have little organizational structure when the z dimension is collapsed.[36] Given the large size of image stacks relative to a single 2D image, engaging in any systematic search strategy that reduces memory load and improves image coverage might lead to better performance. In support of this proposal, making smaller eye movements on the 2D plane and searching in one quadrant of the lung at a time (i.e., saccadic amplitude and eye movement index) predicted better task performance, whereas measures of the rate of movement through depth (i.e., scrolling speed and change in XY/Z) did not. Although caution is warranted when interpreting null results with inconclusive Bayes factors (e.g., change in XY/Z), this pattern of results would be unexpected if drilling was a better strategy primarily because of abrupt motion onset cues.

This study has a number of limitations. First, although this study had a larger sample size than similar studies in the medical image perception literature (on average: 7.73 experts, 5.60 intermediates, and 8.36 novices per study[10]), the sample size was still smaller than ideal for an individual differences study due to the inherent difficulty of collecting large samples of expert radiologists. To address this concern, Bayes factors have been included for each analysis to help distinguish between results with sufficient power and those that will require more evidence. Many of these analyses reached the threshold for sufficient evidence in favor of either the null or alternative hypothesis[44] (Tables 1 and 2). However, some of the critical analyses (e.g., the relationship between experience and saccadic amplitude) will require follow-up studies with larger sample sizes to make a strong conclusion. At minimum, given the relatively large sample size of this study, these results suggest experience-related changes in search behavior are likely much smaller in volumetric images than the effects observed in the previous studies with 2D medical images. In addition, the "ground truth" for abnormality presence or absence is difficult to establish when using real medical images. Although the LIDC database includes every nodule that was marked by at least one of the expert observers rather than expert consensus, it is still possible that some of the less conspicuous nodules were missed by all of the expert observers.[66] In future research, it may be fruitful to replicate these findings using simulated nodules or to use a method of analysis that does not require an independent assessment of ground truth.[67,68] Finally, in future studies, it may be beneficial to investigate expert search behavior in a more clinically valid context. For example, this study had a relatively high abnormality prevalence rate, which may have increased observer fatigue or shifted the observers' decision criterion for marking an abnormality.

This study largely replicated previous findings that drilling is a better strategy for lung nodule detection than scanning when controlling for the effects of experience.[36] However, there are significant challenges in how to classify radiologists as scanners or drillers. In this study, like others in the literature, we initially divided the radiologists into groups by analyzing their depth by time plots and subjectively categorizing them based on the scan patterns [Fig. 5(a)].[36,39] There are clear and significant limitations to using a subjective approach for categorizing search strategy, but there is currently no consensus on how to best capture search strategy using quantitative metrics. In the original scanner/driller study, the subjective categorization method closely matched each observer's eye movement index[30] (see also Ref. 33). Here using quartiles to objectively divide radiologists into groups based on this metric also closely (but imperfectly) matched the subjective groups [Fig. 5(b)] and independently predicted task performance as both a dichotomous [quartile analysis, Fig. 6(b)] and continuous [regression analysis, Fig. 7(a)] variable. Change in XY/Z score also roughly matched the subjective categorization of search strategy but did not match the groups as well as the eye movement index [Fig. 5(c)].[38] Unlike the eye movement index, change in XY/Z did not independently predict performance in this task [Figs. 6(c) and 7(b)], suggesting these metrics may reflect different aspects of search behavior. This study is the first to compare all three methods of characterizing drilling behavior in relation to task performance, and these results suggest the eye movement index might be a suitable alternative to subjective categorization methods. In addition, future studies might use a data-driven approach (e.g., principal components analysis) or instruct observers to use a particular search strategy to determine which eye tracking metrics are able to best classify the groups. In addition, it was previously unclear if it is more appropriate to use quantitative measures to separate radiologists into distinct categories (i.e., the quartile analysis) or as continuous predictors of search behavior (i.e., the regression analysis). Here the quartile analysis and the regression analysis showed the same pattern of results, suggesting it may be unnecessary to divide radiologists into distinct groups. Finally, we still need to establish the reliability of search strategy within an observer and between tasks to determine whether search strategy is internal to a radiologist or unique to specific task circumstances.

The current findings suggest that global processing plays a lesser role in volumetric image interpretation than in 2D analogous tasks, but alternative accounts for these results should be considered. Here we quantified experience in terms of both the overall number of years spent practicing radiology (i.e., years of experience) and the degree of routine experience with the task (i.e., chest CTs per week). However, the number of chest CTs read per week did not relate to any of our key variables, and it is unclear whether self-reported estimates of task experience are

reliable. Critically, however, none of our results substantively differ if only years of experience are included in the regression models. In addition, the eye tracking and behavioral measures associated with global processing ability in 2D medical image interpretation may not tap into the same cognitive process in volumetric image interpretation. For example, saccadic amplitude is highly confounded with search strategy. By definition, scanners have a larger saccadic amplitude than drillers because they engage in larger, sweeping eye movements across the 2D plane. Similarly, it is unclear how to best classify miss errors in volumetric images. Although we categorized search errors in volumetric medical images in the same way as previous studies using 2D images, it is debatable whether the 1000-ms threshold used for distinguishing between recognition and decision errors in 2D images is appropriate for volumetric data. Prolonged nodule fixations might be less common for dynamic stimuli than 2D images, which could artificially inflate the number of recognition errors in volumetric images. In future research, it would be beneficial to use a data-driven approach with a larger stimulus set in order to identify an appropriate threshold.[69] Finally, it is unclear how to best address the inherent differences in the time to first fixation measure between 2D and volumetric images. In volumetric images, radiologists have multiple opportunities to detect an abnormality when scrolling back and forth through depth. As a result, we calculated time to first fixation relative to the moment the abnormality becomes visible during the instance it was detected. Here to avoid some of these concerns, we focused on a set of metrics that would reflect global processing ability rather than focusing heavily on the results of any single metric. Because enhanced global processing ability results in a wider UFOV, one would predict that a global search strategy would be associated with shorter search times, reduced time to first fixation, and smaller image coverage regardless of whether the images are 2D or volumetric. Across each of these measures, we did not find evidence that any of these global processing measures differed with experience in this task. However, as the global properties of volumetric medical images are not yet well-defined, there is ample opportunity for additional research in this area.

This research may ultimately provide insight on how radiology residents should be trained to search through volumetric medical images. In 2D interpretation tasks, translating expertise-related changes in search behavior into training techniques has proven to be quite difficult. Because experts' enhanced perceptual abilities are closely linked to repeated exposure to medical images,[70–72] efforts to train novices to adopt the search patterns of experts have been largely unsuccessful at improving diagnostic accuracy, and there are currently no known "shortcuts" to enhanced global processing ability in radiology.[73–76] Here we did not find experience-related differences in search behavior that might reflect a more global search strategy. Instead, we found that individual differences in task performance were closely related to whether the observer drilled through the image slices and searched the images thoroughly. These results are intriguing because they suggest that instructing radiologists to engage in these search behaviors during training could translate to better diagnostic accuracy in clinical practice.[65]

5 Conclusion

Although research dating back to the early 1970s has demonstrated that experience improves global processing ability, this study is the first to test this prediction in a volumetric image interpretation task while allowing radiologists to freely scroll through the image slices. Across a wide range of measures that have been associated with experience in previous research, we found evidence that experience was not predictive of performance when searching volumetric medical images. These findings suggest the ability to extract the global statistical properties of an image might be more difficult in image stacks. Rather than individual differences in global processing ability, diagnostic performance was closely related to whether radiologists engaged in drilling versus scanning, with drilling being a more thorough, systematic search of the image that resulted in better detection. In future research, it may be fruitful to focus on whether instructing radiologists to use a drilling strategy improves image coverage and task performance. Overall, these findings demonstrate that existing models of perceptual expertise in radiology do not fully account for search behavior in volumetric images, and addressing this gap in the literature is a promising avenue for future research.

Disclosures

The authors declare they have no competing interests.

Acknowledgements

We would like to acknowledge the NCI Perception Laboratory at RSNA and David Alonso for the help with participant recruitment as well as the radiologists that participated in our study. We were also grateful to the NIH (Grant #1R01CA225585-01 for T. D.) and the NSF Graduate Research Fellowship Program (No. #1747505 for L. H. W.).

References

1. L. Berlin, "Accuracy of diagnostic procedures: has it improved over the past five decades?" *Am. J. Roentgenol.* **188**(5), 1173–1178 (2007).
2. T. Donovan and D. Litchfield, "Looking for cancer: expertise related differences in searching and decision making," *Appl. Cognit. Psychol.* **27**, 43–49 (2013).
3. H. L. Kundel et al., "Holistic component of image perception in mammogram interpretation: gaze-tracking study," *Radiology* **242**, 396–402 (2007).
4. H. L. Kundel et al., "Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms," *Acad. Radiol.* **15**, 881–886 (2008).
5. D. P. Carmody, C. F. Nodine, and H. L. Kundel, "Finding lung nodules with and without comparative visual scanning," *Percept. Psychophys.* **29**, 594–598 (1981).
6. A. J. Carrigan, S. G. Wardle, and A. N. Rich, "Finding cancer in mammograms: if you know it's there, do you know where?" *Cognit. Res. Princ. Implic.* **3**(1), 10 (2018).
7. K. K. Evans et al., "The gist of the abnormal: above-chance medical decision making in the blink of an eye," *Psychon. Bull. Rev.* **20**, 1170–1175 (2013).
8. H. L. Kundel and C. F. Nodine, "Interpreting chest radiographs without visual search," *Radiology* **116**, 527–532 (1975).
9. S. Brams et al., "The relationship between gaze behavior, expertise, and performance: a systematic review," *Psychol. Bull.* **145**(10), 980–1027 (2019).
10. A. Gegenfurtner, E. Lehtinen, and R. Säljö, "Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains," *Educ. Psychol. Rev.* **23**, 523–552 (2011).
11. A. van der Gijp et al., "Interpretation of radiological images: towards a framework of knowledge and skills," *Adv. Health Sci. Educ. Theory Pract.* **19**(4), 565–580 (2014).
12. J. W. Oestmann, "Lung lesions: correlation between viewing time and detection," *Radiology* **166**(2), 451–453 (1988).
13. H. L. Kundel and P. S. La Follette, "Visual search patterns and experience with radiological images," *Radiology* **103**(3), 523–528 (1972).
14. B. S. Kelly et al., "The development of expertise in radiology: in chest radiograph interpretation, "expert" search pattern may predate "expert" levels of diagnostic accuracy for pneumothorax identification," *Radiology* **280**(1), 252–260 (2016).
15. T. Drew et al., "Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images?" *Radiographics* **33**, 263–274 (2013).
16. C. F. Nodine and H. L. Kundel, "The cognitive side of visual search in radiology," in *Eye Movements from Physiology to Cognition*, J. K. O'Regan and A. Levy-Schoen, Eds., Elsevier (1987).
17. R. G. Swensson, "A two-stage detection model applied to skilled visual search by radiologists," *Percept. Psychophys.* **27**, 11–16 (1980).
18. J. M. Wolfe et al., "Visual search in scenes involves selective and nonselective pathways," *Trends Cognit. Sci.* **15**(2), 77–84 (2011).
19. S. C. Chong and A. Treisman, "Representation of statistical properties," *Vision Res.* **43**(4), 393–404 (2003).

20. L. Parkes et al., "Compulsory averaging of crowded orientation signals in human vision," *Nat. Neurosci.* **4**(7), 739–744 (2001).
21. M. R. Greene and A. Oliva, "Recognition of natural scenes from global properties: seeing the forest without representing the trees," *Cognit. Psychol.* **58**(2), 137–176 (2009).
22. D. W. Williams and R. Sekuler, "Coherent global motion percepts from stochastic local motions," *ACM SIGGRAPH Comput. Graphics* **18**(1), 24–24 (1984).
23. G. Müller-Plath and K. Elsner, "Space-based and object-based capacity limitations in visual search," *Vis Cognit.* **15**(5), 599–634 (2007).
24. M. R. Greene and A. Oliva, "The briefest of glances: the time course of natural scene understanding," *Psychol. Sci.* **20**(4), 464–472 (2009).
25. M. C. Potter, "Short-term conceptual memory for pictures," *J. Exp. Psychol. [Hum. Learn.]* **2**(5), 509–522 (1976).
26. K. P. Andriole et al., "Optimizing analysis, visualization, and navigation of large image data sets: one 5000-section CT scan can ruin your whole day," *Radiology* **259**, 346–362 (2011).
27. R. J. McDonald et al., "The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload," *Acad Radiol.* **22**, 1191–1198 (2015).
28. L. H. Williams and T. Drew, "What do we know about volumetric medical image interpretation?: a review of the basic science and medical image perception literatures," *Cognit. Res. Princ. Implic.* **4**(1), 21 (2019).
29. M. Treviño et al., "Rapid perceptual processing in two- and three-dimensional prostate images," *J. Med. Imaging* **7**(2), 022406 (2020).
30. C.-C. Wu et al., "Gist processing in digital breast tomosynthesis," *J. Med. Imaging* **7**(2), 022403 (2019).
31. A. Venjakob et al., "Radiologists' eye gaze when reading cranial CT images," *Proc. SPIE* **8318**, 83180B (2012).
32. D. van Montfort et al., "Expertise development in volumetric image interpretation of radiology residents: what do longitudinal scroll data reveal?" *Adv. Health Sci. Educ.* **26**, 1–30 (2020).
33. R. Bertram et al., "Eye movements of radiologists reflect expertise in CT study interpretation: a potential tool to measure resident development," *Radiology* **281**, 805–815 (2016).
34. R. Bertram et al., "The effect of expertise on eye movement behaviour in medical image perception," *PLoS One* **8**, e66169 (2013).
35. I. Diaz et al., "Eye-tracking of nodule detection in lung CT volumetric data," *Med. Phys.* **42**, 2925–2932 (2015).
36. T. Drew et al., "Scanners and drillers: characterizing expert visual search through volumetric images," *J. Vision* **13**, 3 (2013).
37. S. E. Seltzer et al., "Spiral CT of the chest: comparison of cine and film-based viewing," *Radiology* **197**, 73–78 (1995).
38. A. Aizenman et al., "Comparing search patterns in digital breast tomosynthesis and full-field digital mammography: an eye tracking study," *J. Med. Imaging* **4**(4), 045501 (2017).
39. L. C. Kelahan et al., "The radiologist's gaze: mapping three-dimensional visual search in computed tomography of the abdomen and pelvis," *J. Digital Imaging* **32**(2), 234–240 (2019).
40. R. A. Abrams and S. E. Christ, "Motion onset captures attention," *Psychol. Sci.* **14**(5), 427–432 (2003).
41. L. Williams et al., "The invisible breast cancer: experience does not protect against inattentional blindness to clinically relevant findings in radiology," *Psychon. Bull. Rev.* **28**, 503–511 (2020).
42. S. G. Armato et al., "The Lung Image Database Consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," *Med. Phys.* **38**(2), 915–931 (2011).
43. L. Williams et al., *What are the characteristics of expertise in volumetric medical image search?* (2019).
44. H. Jeffreys, *The Theory of Probability*, Oxford University Press, Oxford (1998).
45. H. L. Kundel, C. F. Nodine, and D. Carmody, "Visual scanning, pattern recognition and decision-making in pulmonary nodule detection," *Invest. Radiol.* **13**, 175–181 (1978).

46. H. L. Kundel et al., "Searching for lung nodules. A comparison of human performance with random and systematic scanning models," *Invest. Radiol.* **22**, 417–422 (1987).
47. G. D. Rubin, "Lung nodule and cancer detection in computed tomography screening," *J. Thorac. Imaging* **30**, 130–138 (2015).
48. S. H. Adamo et al., "Mammography to tomosynthesis: examining the differences between two-dimensional and segmented-three-dimensional visual search," *Cognit. Res. Princ. Implic.* **3**, 17 (2018).
49. M. A. Lago et al., "Interactions of lesion detectability and size across single-slice DBT and 3D DBT," *Proc. SPIE* **10577**, 105770X (2018).
50. E. Helbren et al., "Towards a framework for analysis of eye-tracking studies in the three dimensional environment: a study of visual search by experienced readers of endoluminal CT colonography," *Br. J. Radiol.* **87**, 20130614 (2014).
51. L. H. Williams and T. Drew, "Distraction in diagnostic radiology: how is search through volumetric medical images affected by interruptions?" *Cognit. Res. Princ. Implic.* **2**, 12 (2017).
52. A. Ba et al., "Search of low-contrast liver lesions in abdominal CT: the importance of scrolling behavior," *J. Med. Imaging* **7**(4), 045501 (2020).
53. P. C. Brennan et al., "Radiologists can detect the 'gist' of breast cancer before any overt signs of cancer appear," *Sci. Rep.* **8**(1), 1–12 (2018).
54. K. K. Evans et al., "A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast," *Proc. Natl. Acad. Sci. U. S. A.* **113**(37), 10292–10297 (2016).
55. C. A. Dickinson and G. J. Zelinsky, "Memory for the search path: evidence for a high-capacity representation of search history," *Vision Res.* **47**, 1745–1755 (2007).
56. H. J. Godwin, V. Benson, and D. Drieghe, "Using interrupted visual displays to explore the capacity, time course, and format of fixation plans during visual search," *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1700–1712. (2013).
57. R. M. Klein and W. J. MacInnes, "Inhibition of return is a foraging facilitator in visual search," *Psychol. Sci.* **10**, 346–352 (1999).
58. M. S. Peterson, M. R. Beck, and M. Vomela, "Visual search is guided by prospective and retrospective memory," *Percept. Psychophys.* **69**, 123–135 (2007).
59. T. Foulsham and A. Kingstone, "Where have eye been? Observers can recognise their own fixations," *Perception* **42**, 1085–1089 (2013).
60. M. L. H. Võ, A. M. Aizenman, and J. M. Wolfe, "You think you know where you looked? You better look again," *J. Exp. Psychol. Hum. Percept. Perform.* **42**, 1477–1481 (2016).
61. M. Wermeskerken, D. Litchfield, and T. Gog, "What am I looking at? Interpreting dynamic and static gaze displays," *Cognit. Sci.* **42**, 220–252 (2018).
62. T. Drew et al., "Quantifying the costs of interruption during diagnostic radiology interpretation using mobile eye-tracking glasses," *J. Med. Imaging* **5**(3), 031406 (2018).
63. M. M. Alakhras et al., "Effect of radiologists' experience on breast cancer detection and localization using digital breast tomosynthesis," *Eur. Radiol.* **25**, 402–409 (2015).
64. T. Blanchon et al., "Baseline results of the Depiscan study: a French randomized pilot trial of lung cancer screening comparing low dose CT scan (LDCT) and chest X-ray (CXR)," *Lung Cancer* **58**, 50–58 (2007).
65. A. van der Gijp et al., "The effect of teaching search strategies on perceptual performance," *Acad. Radiol.* **24**(6), 762–767 (2017).
66. M. P. Eckstein et al., "Quantifying the limitations of the use of consensus expert committees in ROC studies," *Proc. SPIE* **3340**, 128–134 (1998).
67. R. M. Henkelman, I. Kay, and M. J. Bronskill, "Receiver operator characteristic (ROC) analysis without truth," *Med. Decis. Making* **10**(1), 24–29 (1990).
68. H. L. Kundel and M. Polansky, "Mixture distribution and receiver operating characteristic analysis of bedside chest imaging with screen-film and computed radiography," *Acad. Radiol.* **4**(1), 1–7 (1997).
69. M. S. Cain, S. H. Adamo, and S. R. Mitroff, "A taxonomy of errors in multiple-target visual search," *Vis. Cognit.* **21**, 899–921 (2013).

70. W. Chen et al., "Perceptual training to improve hip fracture identification in conventional radiographs," *PLoS One* **12**, e0189192 (2017).
71. C. Mello-Thoms, "How much agreement is there in the visual search strategy of experts reading mammograms?" *Proc. SPIE* **6917**, 691704 (2008).
72. L. Z. Sha et al., "Perceptual learning in the identification of lung cancer in chest radiographs," *Cognit. Res. Princ. Implic.* **5**(1), 4 (2020).
73. K. Geel et al., "Teaching systematic viewing to final-year medical students improves systematicity but not coverage or detection of radiologic abnormalities," *J. Am. Coll. Radiol.* **14**, 235–241 (2017).
74. A. Gegenfurtner et al., "Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis," *Comput. Educ.* **113**, 212–225 (2017).
75. E. M. Kok et al., "Systematic viewing in radiology: seeing more, missing less?" *Adv. Health Sci. Educ.* **21**, 189–205 (2016).
76. D. Litchfield et al., "Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection," *J. Exp. Psychol. Appl.* **16**, 251–262 (2010).

Biographies of the authors are not available.