REVIEW ARTICLE

Visual search in breast imaging

[1]ZIBA GANDOMKAR, BSc, MSc, PhD and [2]CLAUDIA MELLO-THOMS, BSc, MSc, PhD

[1]BreastScreen Reader Assessment Strategy (BREAST), Discipline of Medical Imaging Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia
[2]Department of Radiology, Carver College of Medicine, University of Iowa, Iowa City, IA, US

Address correspondence to: Prof Claudia Mello-Thoms
E-mail: *claudia-mello-thoms@uiowa.edu*

ABSTRACT

Breast cancer is the most common cancer among females worldwide and large volumes of breast images are produced and interpreted annually. As long as radiologists interpret these images, the diagnostic accuracy will be limited by human factors and both false-positive and false-negative errors might occur. By understanding visual search in breast images, we may be able to identify causes of diagnostic errors, find ways to reduce them, and also provide a better education to radiology residents. Many visual search studies in breast radiology have been devoted to mammography. These studies showed that 70% of missed lesions on mammograms attract radiologists' visual attention and that a plethora of different reasons, such as satisfaction of search, incorrect background sampling, and incorrect first impression can cause diagnostic errors in the interpretation of mammograms. Recently, highly accurate tools, which rely on both eye-tracking data and the content of the mammogram, have been proposed to provide feedback to the radiologists. Improving these tools and determining the optimal pathway to integrate them in the radiology workflow could be a possible line of future research. Moreover, in the past few years deep learning has led to improving diagnostic accuracy of computerized diagnostic tools and visual search studies will be required to understand how radiologists interact with the prompts from these tools, and to identify the best way to utilize them. Visual search in other breast imaging modalities, such as breast ultrasound and digital breast tomosynthesis, have so far received less attention, probably due to associated complexities of eye-tracking monitoring and analysing the data. For example, in digital breast tomosynthesis, scrolling through the image results in longer trials, adds a new factor to the study's complexity and makes calculation of gaze parameters more difficult. However, considering the wide utilization of three-dimensional imaging modalities, more visual search studies involving reading stack-view examinations are required in the future. To conclude, in the past few decades visual search studies provided extensive understanding about underlying reasons for diagnostic errors in breast radiology and characterized differences between experts' and novices' visual search patterns. Further visual search studies are required to investigate radiologists' interaction with relatively newer imaging modalities and artificial intelligence tools.

INTRODUCTION

Visual search plays a crucial role in our daily life and involves easy tasks such as finding our tea cup on a table or difficult ones which require certain level of experience and training. Among the visual search tasks, those conducted by health professionals are among the most complex and differ from other tasks in various aspects.[1] First, the cost of error caused by incorrect medical image interpretation is considerably high compared to our everyday tasks. Secondly, in medical images, the nature of the target and surrounding distractors differ from simple search tasks. When dealing with medical images, the number of targets and target categories might be unknown or unlimited and targets might be hard to find or identify. In some cases, such as screening mammography, low target prevalence might change radiologists' visual search patterns. Moreover, most of the traditional visual search models, which successfully describe simple visual tasks, do not consider observer's knowledge or experience, which are essential in medical image interpretation.[2] Basically, medical image interpretation is comprised of two processes: visual inspection of image (visual perception) and rendering a diagnostic decision. In both processes, efficiency and accuracy highly depend on the observer's expertise.[3]

As breast cancer is the most common cancer among females, a large volume of breast images is interpreted each year. This highlights why visual search studies focused on breast imaging have been necessary. Although most of these studies focused on mammographic image interpretation, visual search in other imaging modalities such as digital breast tomosynthesis (DBT) or breast ultrasound have

also been investigated. Mammography is the primary imaging modality in the investigation of breast changes.[4] Breast ultrasound, DBT, and breast MRI are also widely used to detect breast cancers in high risk females and for complete evaluation of breast symptoms.[4] Other breast imaging modalities include CT, positron emission tomography (PET), electrical impedance-based imaging, scintimammography, and optical imaging.[4] Although, at a fundamental level, the interpretation of all different image modalities shares similarities in terms of involving perceptual and cognitive processes,[3] visual search in each modality has its own unique challenges and properties. This article briefly reviews visual search studies done in different imaging modalities and identifies the questions still to be answered to address gaps in knowledge. To understand medical image interpretation procedures, scientists have widely used eye tracking technology. Therefore, first we briefly describe how eye tracking data is usually analyzed.

Eye tracking technology

Eye tracking technology is an important tool to understand medical image interpretation procedures. Eye trackers are sensors that show exactly where the eyes are focused and measure eye movements. Eye tracking data are often analyzed in terms of fixations and saccades. Fixations represent those times when the observer's eyes pause over an area and they use the central foveal vision to obtain detailed information about the area. Saccades are rapid, often ballistic movements between fixations, which reposition the eyes. An important step done by most of the eye tracking software applications is converting raw eye position data to a sequence of fixations. Raw eye position data contain gaze points representing the instantaneous spatial locations [an (x,y) coordinate] and a timestamp corresponding to each measurement of location. The sampling frequency of the eye tracking device determines how the raw data are spaced temporally. For example, at 500 Hz, time stamps are spaced 2 ms (1/500 s) apart. A fixation contains multiple gaze points and hence has duration, start and end timestamps in addition to a spatial location. There are different algorithms to translate the raw sequence of gaze points into an associated sequence of fixations, but basically they group spatially and temporally adjacent gaze points into a single a fixation. Figure 1 represents a sample eye tracking data, which indicates fixations over a mammogram (red circles) and eye movement trajectories (red lines). The represented data are simulated data.

After collecting eye tracking data, nearby fixations are usually grouped together to generate fixation clusters.[5] This can be done based on spatial proximity (falling within certain distance of the centre, discarding temporal adjacency) of fixations or both temporal and spatial adjacency.[5] Commonly extracted gaze parameters from clusters include total dwell time on each cluster, time to hit the cluster (first fixate on the cluster), number of fixations per cluster, maximum number of consecutive fixations on cluster, and saccade amplitude of the first fixation in a cluster (*i.e.* the angular distance that the eye travels when fixating on a cluster for the first time).[6–9] Total dwell time of each fixation cluster is also shown in Figure 1 (c). Another common way to represent fixation clusters is using eye-tracking focus maps or

heatmaps, which indicate aggregated dwell time in each cluster and are usually projected back onto the original image. As shown in Figure 1 (c), the duration of fixations in the far left (showed using an arrow on 1 (c)) was only 130 ms. That one fixation does not form a "cluster", therefore that little fixation on the left of Figure 1(a) is negligible.

## MAMMOGRAPHY

Findings from previous studies

Mammography is the most common breast imaging modality and large volumes of mammographic images are produced each year.[4] Therefore, several visual search studies have been conducted devoted to characterizing radiologists' mammographic scanpaths in order to provide a deeper insight into an effective search strategy (which will likely lead to correct diagnoses) and to determining the factors that influence the radiological visual search.[2] Success in characterizing visual search patterns that lead to errors can open opportunities for utilizing eye-tracking data to prevent diagnostic errors[7] and improve radiology educational programs.

Modelling radiologists' visual search pattern

At a fundamental level, directed search and free search can be considered as two possible search strategies for interpreting a medical image. Directed search refers to a search strategy that follows a preconceived plan (*e.g.* the radiologist has a specific pattern for investigating a chest radiograph, starting in the left upper lobe of the lungs), while free search is inspecting the image without a preconceived idea of what it may depict. More than four decades ago, early eye-tracking experiments suggested that the orderly approach for visual search instructed in radiology textbooks is not reflected in the eye-movement behaviours observed in practice, and both experienced and less-experienced radiologists use free search rather than directed search.[10] When an image is abnormal, radiologists fixate promptly on the abnormality and neglect relatively large areas of image. Indeed, radiologists can identify abnormal images based on a momentary glimpse of image before detailed fixations at any location.[11] This highlights the importance of the holistic processing of a medical image, which starts immediately after image onset.

Different theoretical frameworks, including the two-state detection model,[12] the global-focal search model,[13] and the holistic model,[14] have included holistic or global processing as a core element characterizing the underlying mechanisms of medical image perception. According to the global-focal search model (Figure 2), immediately after the image onset, the scene is globally analysed with low resolution peripheral vision (generating a "global impression") and the perturbations in the image are identified. Then, gaze is directed to the perturbations, and they are analyzed by using the high-resolution central vision ("fixation checking"). After making a covert decision about the perturbation, the gaze is either directed to another suspicious location based on information from the global impression step or the radiologist starts "discovery scanning" of the image. Discovery scanning may be interrupted by checking fixations if the observer identifies perturbations based on peripheral vision. Meanwhile, the observer may revisit the perturbations

Figure 1. (a) Example of mammographic scanpath (simulated signal); each circle represents a fixation. (b) corresponding clusters of fixations based on spatial information and discarding temporal information; each colour represents centre of one of the clusters. Clusters are randomly colour-coded in a way to ensure that neighbouring clusters do not have similar colours. The original fixation points were plotted in light grey. (c) Corresponding clusters of fixations based on spatial and temporal information; the centre of clusters are randomly colour-coded. The numbers indicate the total dwell time of each fixation cluster. As indicated, the cluster shown using an arrow had the shortest duration (0.13 s). (d) Corresponding eye-tracking focus map (the intensity of green colour shows the total dwell time on an area).
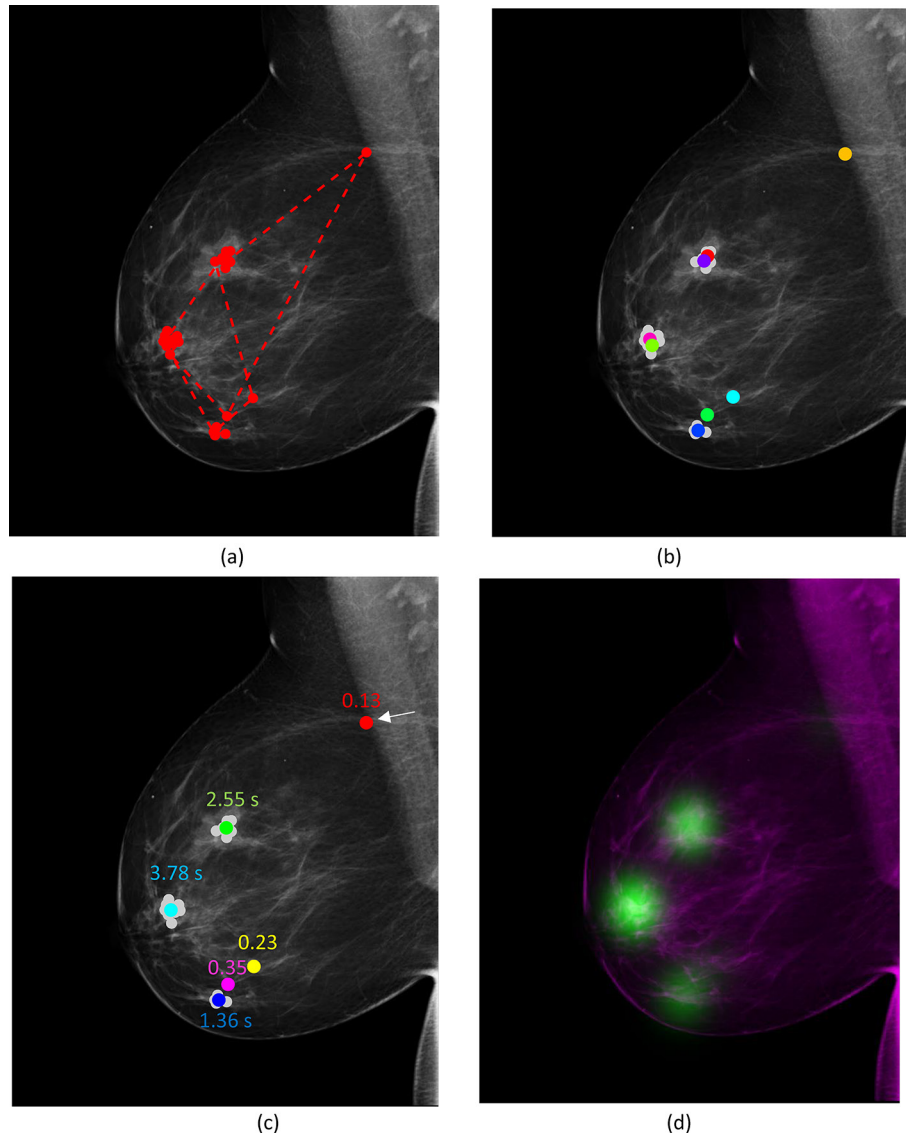

several times to compare them with normal tissue or to make difficult decisions about ambiguous features in the image. Although the global-focal search model initially described visual search for pulmonary nodules,[15] later studies indicated its success in describing mammographic visual search.[14,16] For example, the distribution of time to hit the lesion (indicator of the initial perception of cancer) was estimated by a mixture of two Gaussian distributions, the first one representing lesion perception happening as a result of the global impression and the second one representing lesion discovery during subsequent fixations.[17] It has been shown that the true-positive fraction of the first group (the lesions hit shortly after image onset) was large

for most readers, and concluded that this initial detection occurs as a result of parallel global analysis of the image.[14,17]

The two-stage detection model[12] is very similar to the global-focal search model. According to this model, expertise development in medical image perception involves developing an initial pre-attentive filter that automatically detects features that need further inspection and filtering out normal anatomical structures, in order to rapidly direct the expert's attention toward regions of the image that contain probable abnormalities. Thus, both global-focal and two-stage detection models suggest that experts can process large areas of an image using their peripheral vision

Figure 2. Steps of global-focal search model. The figure is adopted from the diagram provided in Nodine and Kundel.[13] As shown, visual search is modelled as a serial process, which starts with capturing the global impression which is based on prior knowledge. Once the perturbations in the image are identified (as shown by the rightmost arrows), they can be either verified foveally, or used as input for Discovery or Reflective Search to specify a set of potential areas or features to be fixated. Information collected by the detailed inspection using the foveal vision combined with prior knowledge can lead to the final decision for a case. Discovery Search refers to coarse screening of the image in order to detect potential target features, while Reflective Search aims at building up enough visual evidence to make a decision on an ambiguous target through detailed image screening. In experimental conditions where images are presented for a limited time, radiologists are required to make their final decision when the image is no longer presented based on "Post-Search Recall". As shown, Global and Focal steps can be done recursively.
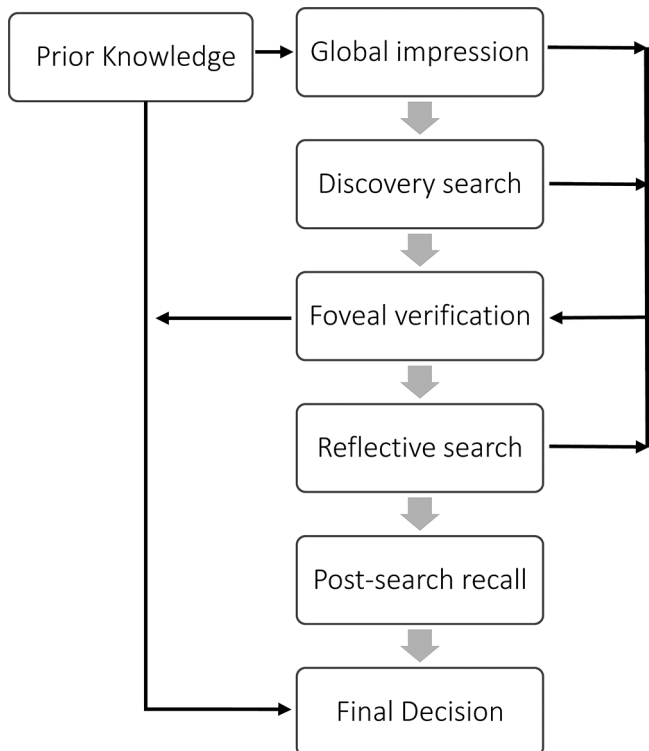
and rapidly identify potential abnormal areas, which should be subsequently inspected using foveal vision. The main difference between the global-focal search model and the two-stage detection model is the assumption that the two stages (*i.e.* global and local processing) can be done recursively.

As presented in Figure 3, the holistic model is concordant with the two-stage detection and global-focal search models in terms of including holistic and local processing pathways. It suggests that expertise development involves a shift in the perception mechanism from a relatively slow search-to-find mode to a fast holistic mode. Unlike the two earlier models, which assume that global and local processing are two serial stages, the holistic model assumes that the holistic processing can work in parallel

with the search-to-find mode. This aligns with several other models, trying to conceptualize the holistic processing in other visual search domains,[18] including the view from the basic visual science, where visual search is explained by a selective and a non-selective pathways. Based on these models, in the non-selective pathway, information can be extracted from global image properties.[19]

The holistic model assumes that radiologists' first global impression, or "gist" signal, arising from the mammogram, has a local source of information within the image, *i.e.* a perturbation or an abnormality. However, recent findings questioned this assumption. First, Evans et al[20] showed that after a half-second presentation of mammograms, radiologists can distinguish normal from abnormal at above-chance levels but they perform at chance-level in the abnormality localization. More recently Evans et al suggested that radiologists can identify abnormal cases from normal ones based on a half-second glimpse of normal breasts contralateral to malignancies.[21] Moreover, radiologists performed at above-chance level in differentiating normal prior mammograms belonging to females who were diagnosed with breast cancer in a subsequent screening.[22] The fact that both prior mammograms and contralateral mammograms contain no apparent visible signs of cancer suggests that radiologists' first impression might not have a localized source, presenting a perturbation in the image, but depend on global image statistics suggestive of an abnormal mammographic texture.[20–22] The importance of these new findings about the gist signal is two-fold. First, these findings suggest that modifying the holistic processing model might be required, as it currently assumes that holistic processing provides scene structural information that directs eye movement to areas that could contain items of interest. However, recent experiments showed holistic processing might also involve a global process, which roughly classifies images as, *e.g.* cancer or normal.[23] Secondly, the gist signal can be used to predict a subsequent breast cancer incidence.[24]

Reader's characteristics and visual search
Although models for medical image interpretation aim at characterizing mammographic scanpaths, eye tracking of radiologists indicates differences in scanpaths and areas that attracted viewers' attention. Extensive studies have been undertaken to understand how reader's characteristics, especially their expertise level, affect mammographic scanpaths.

The holistic model characterizes expertise development as a shift in the recognition mechanism from "scan-look-detect" to "look-detect-scan".[17] This shift makes experts' scanpaths more efficient as they are better at using the peripheral vision to process large image areas in parallel. Figure 4 (a) , (b) show two fixation sequences, one corresponding to a highly experienced radiologist and one corresponding to a junior radiologist, who just completed breast imaging fellowship. As shown, the two sequences differ considerably. The holistic model can successfully explain several observed differences between the experts' mammographic scanpaths and those of less-experienced readers. Based on this model, the experts should fixate on the lesion earlier as they first "detect" the abnormality and then "scan" the remaining of the image. Based on the holistic model,

Figure 3. Steps of the holistic processing model. As shown, visual search starts with capturing the global impression and it is followed by two parallel processes.
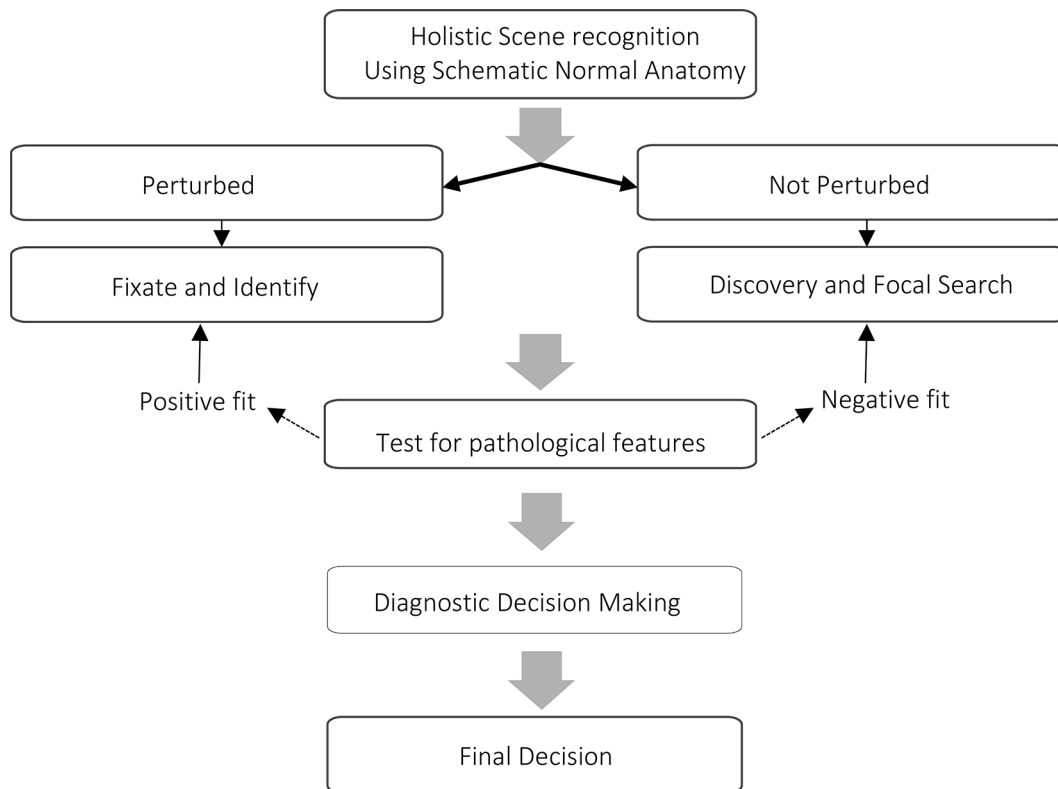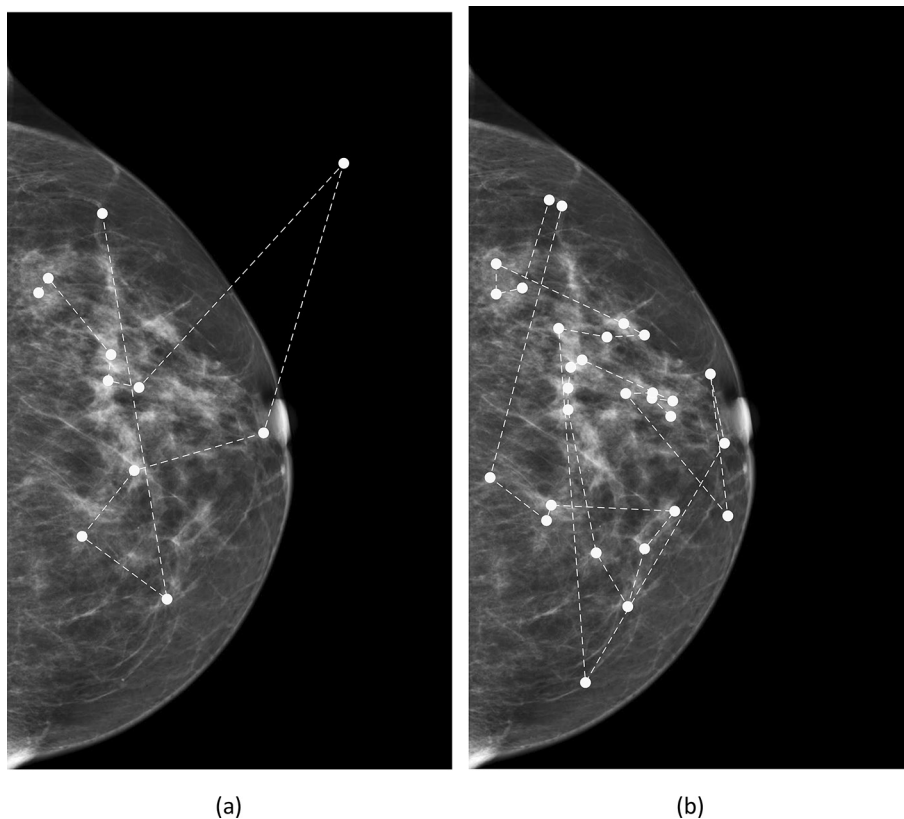

Figure 4. Fixation sequences of (a) a highly experienced radiologist and (b) a junior radiologist, while inspecting a normal case.

experts developed skills process large areas of a mammogram using their peripheral vision, therefore the model expects that experts are more likely to detect potential abnormalities and first fixate on lesions quicker with less image coverage compared to less-experienced readers.

Congruent with the holistic model, compared to less-experienced readers, experts are more accurate[25] while spending less time on each image before making their final decision,[16] first fixate on the lesion earlier during their search,[26] cover relatively smaller areas of the image with foveal fixations,[26] have fewer fixations with longer saccades,[27] and have a smaller number of fixations per saccades.[26] On average, experts have longer dwell times on false-negatives (FNs) while inexperienced observers have longer dwell times on true-positives (TPs), false-positives (FPs), and true-negatives (TNs).[28] Analysis of the speed–accuracy relationship showed that for the least experienced group, overall performance decreased below the chance level when total decision-making time was above 60 s.[29] On average, experts have fewer comparison scans between left and right mammograms compared to inexperienced readers.[30] Experienced radiologists also refixate less often on a previously fixated location and mostly use refixating, laminar (consecutive fixations in a small neighbourhood representing detailed scanning of an area within the image) and deterministic eye movements (repeated short sequences of fixations, typically looking back and forth between two locations for comparison) in the areas containing lesions, thus yielding a more efficient search compared to less experienced readers.[31] Experts' scanpaths show a trend toward a small number of diagnostically relevant areas while less-experienced readers' scanpaths are more random and chaotic.[31] Readers' level of experience is also an independent factor in predicting complexity of visual search behaviour as measured by fractal dimension of mammographic scanpaths.[32] Experts show that a greater proportion of time is spent fixating on diagnostically relevant areas,[33] and greater consistency in their scanpaths in terms of fixated areas.[34] Moreover, a spatial frequency analysis of fixated regions indicated that less experienced readers' search strategies are more significantly affected by the local saliency of the lesions[35] when compared to those of experts.

When compared with radiologists, in addition to experience, a naïve observer lacks training in mammogram interpretation. Nodine et al[26] explored whether training alone, without extensive experience, can yield the experts' efficient visual search strategies. Their results suggested that even after training (i.e. one or more mammography rotations), less experienced readers lacked the ability to efficiently dismiss normal areas with lesion-like features and failed to confirm the presence of a lesion in the second view of the mammogram.[26] It was suggested in a later study[29] that fine-tuned visual recognition skills of radiology trainees might be developed by perceptual-learning, where mentor-guided immediate feedback during image interpretation is provided to trainees.

Showing responses from computer-aided detection (CAD) tools on mammograms could change radiologists' visual search strategies, as CAD's input could direct radiologists' attention to certain suspicious areas annotated by CAD. Krupinski[28] showed that experts and less-experienced readers interact differently with input from CAD tools. For example, the search strategy of less-experienced observers seemed to be different if CAD was available, while the experts searched the image as if there were no CAD prompts. These findings suggested that experts spent more time searching the mammogram before inspecting the area marked by the CAD system. On the other hand, the dwell time on FNs increased after inspecting CAD responses for less experienced observers and became longer than that of experts.[28]

Further studies of radiologists' eye tracking data suggested that there are individual aspects associated with mammographic scanpaths that are not explained simply by differences in expertise level. Alamudun et al[32] showed that the individual differences among radiologist are independent factors in predicting complexity of visual search behaviour as measured by fractal dimension. Also, in Gandomkar et al,[31] it was shown that each reader's scanpath exhibits a unique spatiotemporal pattern that can be successfully differentiated from other readers' scanpaths by feeding quantitative gaze-related parameters to a machine-learning classifier. When a reader was more experienced, the accuracy of the classifier in differentiating his/her scanpaths from others' scanpaths was substantially higher compared to when a less-experienced reader was considered. This implies that experienced radiologists more consistently exhibit their unique spatiotemporal dynamics across all images compared to less-experienced radiologists.

Case properties and visual search pattern
Alamudun et al[32] showed that breast density and case pathology (normal, benign, or malignant) are independent predictors of visual search complexity. Also, the spatiotemporal dynamics of fixations, as measured by recurrence quantification analysis, depend on the breast density and presence of a cancer.[31] The observed dependency was stronger for the experienced radiologists who had a more efficient visual search and were more responsive to the presence of the target (i.e. dense tissue and lesion).[31] This suggests that the extent of the effects of case properties on the visual search pattern of radiologists could depend on readers' characteristics. For example, lesion conspicuity[36,37] and subtlety[38] are associated with significantly different visual search parameters than obvious lesions. Taken together these findings suggest that technological advances to improve lesion visibility and supress dense tissue may have potential in reducing detection errors.

Relationship between visual search and errors in mammogram interpretation
Interpretation of mammograms is a challenging task and subject to both FN and FP errors. Nodine and Kundel[13] divided FN errors into three sub types: search errors (lesions were not fixated at all during visual search), recognition errors (lesions attracted radiologist's attention to some extent but failed to exceed the minimum time needed for perception of an abnormality) and decision-making errors (lesions attracted prolonged attention but were finally interpreted as a normal or benign structures).

The distribution of these different types of errors depends on the reader's experience level. The percentage of search, recognition, and decision-making errors made by experts in the interpretation of mammograms was 24%, 24%, and 52%, respectively, while corresponding percentages were 29%, 42%, and 29% for radiology trainees.[28] Recognition error occurs in the perception step and could be associated with many psychophysiological factors such as lesion's conspicuity.[36,37] Decision-making errors also relate to various factors like lack of expertise,[25] bias from the first fixation[39] or incorrect background sampling.[40]

Beside the classification of errors, eye tracking technology helped in understanding underlying causes of diagnostic errors and the characteristics of visual search patterns, which relate to a greater likelihood of making errors while reading mammograms.[41] One of these characteristics is the total time to make a decision on a case. Using an eye-tracker, Nodine et al[42] analyzed the time course of searching mammograms for suspicious lesions. The analysis showed that TPs were detected faster than FPs during the rapid-phase of visual search (the initial 25 s of case reading for breast radiologists and 40 s for trainees) but after that the FPs continue to increase relative to the TPs. The findings suggested that observers should terminate the visual search when they no longer make a high-confidence decision, as prolonging the search at this point has a low probability of yielding a TP and results in FPs.[42]

The characteristics of background samples, which a reader uses for carrying out comparisons and deciding whether an area is cancer-containing or not, can also predict the diagnostic outcome. Comparing radiologists' fixations while making correct and incorrect decisions revealed the effect of faulty background sampling on likelihood of making an error.[40] In Mello-Thoms,[40] the characteristics of the fixated background areas were extracted using spatial frequency analysis after the observers' eyes hit for the first time the location of TP, FN, and FP areas. For TP areas, the spatial frequency representation of background areas sampled before and after first fixating on the location of the decision did not differ, while for FN and FP areas, significant differences were reported.[40] The analysis showed that after fixating on FN areas for the first time, the correlation of the background samples with the lesion increased while the correlation with the average background decreased. This suggested that the observers were actively trying to insert the lesion in the context of the normal background breast tissue and reconcile the perception of the lesion with its background.[40] For FP areas, the perception of a finding effectively biased further visual search as the observer tried to collect evidence to support the presence of a malignant mass in a lesion-free area of the parenchyma. Therefore, it can be hypothesized that in some of the decision making errors, an incorrect background sampling strategy led to the failure in reporting the lesion.[40]

Studying radiologists' fixations also showed that the radiologist's first decision, which mainly occurs as a result of the global impression, affects the rest of performance.[39] When an observer makes a FP decision according to his/her first impression from the image, then the probability that the radiologist will make a faulty decision as a subsequent outcome is higher than the probability of reporting a true lesion. Therefore, an incorrect initial decision (or an incorrect initial "holistic" view) makes the observer effectively "blind" to the characteristics of the real lesion present.[39]

Satisfaction of search (SOS) can also lead to FN errors. SOS happens when an abnormality is missed because the observer has already detected another abnormality in the same image. Most of the previous studies that explored SOS error and its causes focused on chest images. Earlier studies on chest X-ray and trauma radiography showed SOS could lead to a reduction in the number of TPs while search behaviour, in terms of total inspection time and dwell time in the location of reported abnormality remained unchanged.[43,44] Later, Samuel et al,[45] categorized SOS errors that occurred when finding nodules on chest X-rays into scanning, recognition, and decision-making errors. Their results indicated that SOS was not caused by scanning errors (*i.e.* inadequate scanning of the image), but rather it mostly occurred because the missed nodules did not received prolonged attention (*i.e.* they were recognition errors). In mammography, Mello-Thoms et al[46] showed that detection of abnormalities in a case is greatly associated with their similarities to the first perceived abnormality reported by the radiologist (regardless of being TP or FP). Therefore, like those in chest radiographs and trauma radiography, the study supported the idea that SOS is due to perceptual suppression of recognition instead of faulty scanning or visual neglect of certain regions.[46]

Providing feedback to radiologists based on their search pattern

As stated, studies suggested that more than 70% of missed lesions in mammography attracted the radiologists' attention to some extent (*i.e.* they were recognition and decision-making errors)[47] and time to first hit the location of FP areas differed significantly from the time to hit TP areas.[42] Therefore, early studies suggested that giving feedback based on the observer's gaze pattern might reduce errors.[47] However, in the case of mammograms, there was no significant improvement of reader's performance after providing feedback based on the total dwell time of fixated areas.[48] Later, as discussed next, studies combined total dwell time with other gaze and image-related features and applied machine learning techniques to predict different types of errors.

Voisin et al[9] developed a machine learning-based model to predict diagnostic errors by merging radiologists' gaze behaviour and image characteristics when investigating a region of interest (ROI) containing either a benign or malignant mass. The main limitation of this model was considering an image ROI instead of a whole mammogram. Tourassi et al[8] utilized a more realistic experimental setup, where radiologists viewed the whole mammogram, and image features along with gaze parameters were extracted from each fixated location. However, they used analogue (digitized) images and only included a limited number of cases. In a recent study,[7] an individualized tool for identification of interpretation errors in mammograms, called eye-Computer Assisted Perception, was developed and validated on a relatively large data set containing eye tracking data from eight breast radiologists reading 120 two-view digital mammograms.

Eye-Computer Assisted Perception was personalized for each reader as it was shown each radiologist had his/her own unique error-making pattern. Despite of the promising results obtained in the study, the fact that experimental set-up was different from the real clinical practice (*e.g.* presenting two views of a single breast instead of four views of both breasts and lack of prior images) limits the application of the proposed model in practice.

What's to come

Since the 1990s, mammography has been the frontline screening tool for early detection of breast cancer. As discussed in the previous section, in the past four decades, visual search studies provided a valuable understanding of the mammographic interpretation process and revealed several underlying causes of diagnostic errors. However, the current literature has a few limitations, which can be addressed in future studies on visual search in mammography. First, the experimental design of the previous studies was not consistent with clinical practice. In most of them, using image viewing features such as zooming, inverting, Tabár's systematic viewing mask techniques,[49] and other types of image manipulation were not allowed, as analysis of the eye tracking data can be significantly more difficult when these manipulations are used. However, in the real clinical practice those options are available for radiologists and are frequently used by some individuals.

Moreover, due to difficulty of eye tracking across multiple displays, usually studies are based on a single view mammogram, while in the real practice four views along with prior mammograms (if available) are presented simultaneously. A few previous studies[6,7,23,24,30,31,50,51] used craniocaudal (CC) and mediolateral (MLO) views, which were presented on two displays. When two views were available, radiologists were able to compare between them to make decisions to report perceived findings, and this could result in different mammographic scanpaths as compared to the single mammogram viewing condition. Study of spatio-temporal dynamics of radiologists when viewing CC and MLO views simultaneously showed that radiologist's gaze frequently moves from one view to another, and experts are better in fusing image information from different views.[31] Such a difference would not be observed in a study based on a single view mammogram. Also, based on analysis done,[31] the temporal dynamics of experts' fixations when the cancer is present only on a single view differ from cases where the cancer is visible on two views. In addition, experienced radiologists more frequently move back and forth between diagnostically relevant areas on two views.[31] Although parameters such as time to first hit the lesion or the role of holistic processing will be similar in the two experimental conditions, without recording eye movements on two views, the above-mentioned differences cannot be characterized. Similarly, eye tracking data based on both sides of the breasts will be required to explore how radiologists assess asymmetry-related features and inclusion of prior images will be needed to characterize mammographic scanpaths when a radiologist compares prior and current mammograms. Only in one previous study,[32] where fractal dimension was used to analyse the complexity of eye-tracking data, an experimental set-up with simultaneous viewing of four co-ordinated breast views as typically done in clinical practice was used. When different hanging protocol are allowed, analyzing data could be challenging as radiologists might switch between hanging protocols. In Alamudun et al,[32] the fixations were aggregated back into the predominant hanging protocols used by the reader for each case. Future studies can use a similar strategy to analyse the fixation data when recording eye tracking across multiple views.

Also, most of the earlier studies focused the association of observers' characteristics and visual search patterns.[16,17,25–32] More recent studies showed that radiologists' visual search patterns relate to breast density and case pathology,[31,32] however, further studies are required to determine how visual search relates to different types of malignancies, lesion difficulty, and other features describing image content. Findings from such studies might help in providing a more effective education to trainees and in determining the optimal practice (search strategy) to mitigate diagnostic errors in detecting more difficult lesions, which are missed more frequently.

In recent years, application of deep learning algorithms led to a new generation of computerized tools for providing a second opinion to radiologists.[52,53] Previously, Krupinski[28] studied interaction of both experienced and less-experienced radiologists with CAD prompts. However, previous generations of CAD tools, which relied on classical machine learning framework and hand-crafted features, annotated many FPs on mammograms and were not popular among radiologists.[54] Given that deep learning provides a new perspective for computerized tools, it is necessary to investigate how displaying decisions made by these new, more-accurate computerized tools, changes the radiologists' visual search. Understanding the interaction between the radiologists and the machine's output will ultimately help in optimizing the way that computerized tools should be implemented in the radiological workflow. Beside deep learning-based tools, computerized tools could include computer-aided perception tools,[7] which combine gaze-related features with image features to predict errors. Current studies[6–9] have assumed that radiologist would accept decisions made by computer-aided perception tools. However, in practice these tools would provide feedback to radiologists, who would still make the final decision. A potential future work is exploring how radiologists interact with the output of computer-aided perception tools.

Moreover, even though radiologists have their own unique error making patterns and a personalized model to predict diagnostic errors outperformed a one-size-fits-all model,[8,9,27,30,31] further studies are required to find out how we can tailor educational materials based on each reader's unique error making and visual search patterns,[6] and to investigate any changes to the reader's visual search pattern after providing such educational materials. Despite advancements in technology, the accuracy of mammographic image interpretation still suffers from intrinsic human errors. Improving radiology education through innovative tools, which audit reader's visual search and provide immediate feedback on their errors, may potentially improve diagnostic efficacy.

BREAST ULTRASOUND
Findings from previous studies

Ultrasound is the most common complement to mammography.[55] As with mammography, the observer's capability of recognizing normal tissue appearances and differentiating them from subtle abnormalities is crucial to make an accurate diagnosis.[4] Unlike mammography, where the entire breast area is visible, ultrasound examination only shows a portion of the breast containing the area of interest. Moreover, sometimes the scans are not performed by radiologists, and sonographers (technologists, non-radiologists) perform the diagnostic scan, select images which capture pathologies and present them to the radiologist for the final reporting.[56] Carrigan et al[57] explored the gaze behaviour of breast sonographers and showed that, similar to what was found in studies on mammograms, sonographers committed search, recognition and decision making errors. For example, prolonged fixations were observed on the FN locations.[57] As sonographers' decisions are sometimes made on unrecorded image data, the findings of this study suggested that areas that attracted prolonged sonographer's attention should be presented to radiologists as they might contain a subtle finding, incorrectly recognized as normal tissue by sonographers.

What's to come

In past two decades, breast ultrasound has been widely used in addition to mammography to better characterize suspicious lesions and determine if an ultrasound-guided interventional procedure should be advised for the confirmation of diagnosis. Interpretation of breast ultrasound is a unique perceptual task from different aspects, especially as it requires hand-eye coordination and real-time decision making.[58] Therefore, findings from visual search studies on mammography might not be directly translatable to breast ultrasound and further investigations in this field are required. First, visual search behaviour of radiologists in breast ultrasound have not been studied yet, as the study conducted by Carrigan et al[57] only included breast sonographers. It is unknown whether the visual search patterns of radiologists differ from those of sonographers. Also, the effect of clinical history and presence of abnormality on mammograms may affect the radiologists' visual search behaviour while reading breast ultrasound, but this effect, if it exists, has not been determined. Mostly, ultrasound is used to evaluate a suspicious area found by mammography and hence during its interpretation, radiologists are informed about the location of the area of concern and its appearance on the mammograms.[4] However, for the investigation of breast symptoms in high-risk females under 35 years of age, ultrasound may be the primary (and only) imaging modality.[59]

VOLUMETRIC BREAST IMAGES
Findings from previous studies

Although it is estimated that 25% of the reduction in breast cancer mortality rate in the past decade is a result of mammography screening,[60] mammography has been recently criticized for its relatively low sensitivity and specificity.[61] DBT is an emerging imaging modality with potential to improve cancer detection rate in the screening program.[62] DBT produces a CT-like stack of two-dimensional (2D) images that radiologists can scroll through in depth. When a suspicious area is found, radiologists can examine it in different slices with less masking effect of overlapping fibroglandular tissue.[62] One of the main obstacles in the wide adoption of DBT is the increased time required to interpret a screening examination, as a DBT exam can contain up to 80 images for the reader to examine, and radiologists are facing challenges to adapt their search strategy to this new imaging modality.[63,64] Better understanding of the optimal search strategy for DBT using eye tracking can help readers to develop more efficient interpretation approaches.

Timberg et al[65] eye-tracked 4 observers while reading 55 DBT images, in which simulated lesions were inserted. Readers' performance in free scroll volume browsing was compared to viewing of an initial cine loop, followed by additional free scrolling. The cine loop was produced in fast, medium, and slow frame rates.[65] Based on this study, cine loop with fast and medium frame rates and horizontal presentation speeded the process of selecting relevant suspicious regions in free scrolling.[65] However, the scanpaths were not recorded in this study and having only four observers and simulated lesions limit generalizability of their findings. In another study, Jiang et al[66] eye-tracked three medical physicists while reading DBT, however, the study investigated the correlation between fixation times and sets of morphological feature values extracted from the images. Although the findings have the potential to improve visual search model observers in the detection-localization task of DBT images, the study did not analyse the radiologists' scanpaths and eye movements.

Two recent eye-tracking studies by Aizenman et al[63] and Dong et al[64] studied the visual search behaviour of radiologists while reading DBT in more detailed. Both studies showed that total search duration was longer for DBT, and dwell time in the fixated areas was longer. Aizenman et al[63] indicated that DBT was associated with a saccadic amplitude similar to that of mammography. They also investigated whether DBT observers align with "driller" or "scanner" strategies, previously identified when radiologists searched for lung nodules in volumetric lung CTs.[67] However, there are two main differences between lung CT and DBT images that make comparing these two modalities challenging. First, the number of slices in CT images is larger than DBT. Secondly, image size is smaller in CT images. Therefore, normalized distances in depth and *XY* plane were used. The results suggested that the DBT observers cannot be fully categorized as previously identified "drillers" or "scanners". They drill through the depth of the breast comparatively quickly and cover more depth levels in a 5 s window when compared with "drillers" in the lung CT study. However, unlike the "drillers", the DBT viewers do not fixate roughly in the same area in *XY* plane while flying through the depth. Having said this, the observed movement in *XY* plane was smaller than what observed among "scanners" in Drew et al.[67] The study concluded that the search strategies of DBT observers were more aligned with drilling, albeit it was not exactly like it. In the other study, Dong et al[64] eye-tracked experienced radiologists while presenting them a 2D overview and the DBT images. It was found that the radiologists were inclined to do a global-focal scan over the 2D view to detect the abnormalities and then drilled through the DBT slices to

verify their detections. No significant difference in viewing time between normal and abnormal cases was found.

What's to come

Considering the inherent limitations of mammography due to masking effect as a result of projecting a three-dimensional (3D) area into a 2D plane and technological advancements, volumetric breast imaging has become more widely used in recent years. Given that 3D breast tomograms and 2D mammograms are very different, results of visual search studies in one modality are not directly translatable to the other modality. Generally, due to challenges associated with eye tracking methodology in volumetric images, the number of visual search studies that looked at volumetric images lags behind conventional 2D radiography.[68] For example, in DBT tomograms radiologists can freely scroll through the image. Scrolling back and forth adds an additional factor to the study. To deal with this, some studies presented a video playback of flying through a stack of image slices, rather than individual slices. Furthermore, analyzing the raw eye tracking data, which relies on detecting fixations and saccades, is fundamentally difficult when dealing with the videos or during scrolling because the content of foveated areas may change during the course of one fixation, and the eyes also frequently perform smooth pursuit movements. Several strategies have been proposed to tackle these challenges and adjust the definition of eye tracking parameters for volumetric images.[68]

As mentioned in the previous section, limited visual search studies have been conducted using DBT images, while reading other types of volumetric breast images (*e.g.* breast MRI or CT or functional imaging) have not been studied yet and future studies in this area are required. These imaging modalities can be used in various situations, *e.g.* breast MRI might be used for screening of high-risk females,[69] determining the extent of breast cancer,[70] or distinguishing fibrosis from tumour recurrence in females who underwent breast-conserving surgeries.[71] Depending on the task at hand, the visual search behaviour of the observers might differ. Better understanding of differences and optimal search behaviour can help in training radiologists and increasing the specificity of breast MRI, which is an ongoing area of investigation.[72] Moreover, the radiologist's visual search patterns when using workstations that provide multimodality image viewing has not been studied. Currently, in clinical practice, using multimodality image viewing is gaining popularity and understanding how an experienced radiologist switches between modalities and relates findings in different modalities could help in establishing best practice approaches. Previously, recurrence quantification measures have been used to explain how radiologists switch between views in mammograms and compare two locations in mammographic images.[31] Similar analysis can be done to quantify how they search two different modalities simultaneously.

HOW BASIC VISUAL SEARCH RESEARCH AIDS VISUAL SEARCH STUDIES IN BREAST IMAGING

In addition to studies conducted using breast images, findings from ongoing basic vision science research, which utilized simplified visual search experiments (*e.g.* finding T-shaped targets among L-shaped distractors), are helpful in studying some of the challenges that radiologists face in breast cancer detection. One of the main challenges, particularly in the interpretation of screening mammograms, is the low prevalence of cancer cases, which could reduce the observer's detection rate. To study the effect of low prevalence target on observer's performance, Wolfe et al[73] designed an experiment where observers searched for a target among items drawn at random from other categories. Three different trials with different target frequency (1%, 10% or 50%) were generated. At prevalence of 50%, observers missed 7% of targets, while error rates increased dramatically with decreased prevalence: a prevalence of 1% produced an error rate of 30%. Interesting, observers who participated in the 50% condition before the 1% condition had an average error rate of 26% in 1% condition, while observers who participated in the 1% condition first averaged 46% of errors. This suggests that a training regimen with high prevalence targets before image interpretation might improve the detection rate, as it might counteract the low prevalence effect.[74] In the context of screening mammography, cancer prevalence is only about 0.5%,[75] which is even lower than lowest-prevalence rate in Wolfe et al's study.[73] Certain abnormalities, such as architectural distortions, only appear on less than 0.05% of cases. Mitroff and Biggs[76] called these targets "ultra-rare" items and showed that the error rate was considerably higher for targets with prevalence below 1% than for targets with prevalence above 1%. They also reported a logarithmic relationship between target detection and target frequency.

Another area where taking cognitive laboratory approaches was helpful was to better understand underlying causes of errors is SOS. Fleck et al[77] designed experiments with simplified search displays and showed SOS outside the discipline of radiology for the first time. Their experiments demonstrated that SOS occurs as a result of biased expectations about the low likelihood of specific targets or events and external pressure (*e.g.* time or reward) to do tasks efficiently. Later, Cain et al[78] used an almost similar experimental protocol and investigated the breakdown of SOS errors. About 24% of SOS errors were due to resource depletion, where searchers refixated on a detected high-salience target. The large number of fixations on this target might disrupt the memory for searched locations[79] or it could bias visual search toward items that match the content of working memory.[80] The study also explored the magnitude of "satisfaction" errors, in which searchers terminated their search early. This error was called "strategy error" and only accounted for 6% of the SOS errors. This echoes findings from earlier research done using medical images, where it was reported that SOS errors were not due to insufficient time-on-task. However, an interesting finding about this category of error was that two given participants were mainly committing this type of error. Thus, although overall strategy errors are not the main cause of SOS, the magnitude of strategy errors accounting for SOS in certain individuals might differ considerably from that of the average observer. In a recent study,[81] where individual differences of observers have been taken into account, evidence for the "satisfaction" account of SOS was provided. Specially, a significant and negative correlation between time spent searching after finding a high-salience target and SOS error rates was found. This implies that observers

who are faster to terminate search after detecting the first target are more likely to commit SOS errors.

A new challenge in breast radiology is moving from 2D mammography to 3D tomosynthesis. As mentioned in "Findings from previous studies", there are a few studies that compared the search strategies in these two modalities.[63–66] However, the relatively longer interpretation time of tomosynthesis, the difficulty in accessing radiologists as participants in research studies (due to their time constraints), and the lower feasibility of experimental control when using medical images makes studying the cognitive processes underlying tomosynthesis visual search difficult. In a recent study, Adamo et al[82] proposed an alternative experimental paradigm to explore 3D visual search with both radiologists and non-radiologists, and obtained insight about how radiologists would search tomosynthesis. They used a simplified experiment where observers were asked to find one T-shaped target among L-shaped distractors. The search items were randomly placed within a $15 \times 15 \times 15$ location matrix. The 3D images were either compressed into a single flat plane (analogous to mammograms) or were divided into 30 different slices (analogous to tomosynthesis). Participants were able to scroll through the slices. The performance of professionals (radiology residents and certified radiologists) and non-professionals (undergraduate students) observers were studied in this experiment. The results showed that the performance of the non-professionals matched the performance of professionals and the performance reported in previous radiological studies. Therefore, the study suggests that their experimental paradigm is a cost-effective tool to conduct experiments using non-professional observers for examining relevant visual search performance in this task. Considering the difficulties in recruiting radiologists in visual search studies, such experimental paradigms are important to formulate and test initial hypotheses regarding different aspects of medical image perception.

Beside experimental paradigms in basic visual search, various techniques, which have been proposed in basic vision sciences to analyze eye-tracking data, can be used in medical visual search studies. Among these techniques, clustering the fixations and extracting gaze features from each cluster is probably the most popular technique among medical image perception studies. However, depending on the study's hypotheses and specific aims, other techniques have been also used in new studies. For example, Anderson et al[83] first proposed a framework based on recurrence quantification analysis to study the global and local spatiotemporal characteristics of fixation sequences of observers while looking at different scenes. This framework has been later used in[31] to study how spatiotemporal characteristics radiologists' fixations, more specifically, how they refixate on a previously fixated area, how they compare different areas with each other, and how the temporal dynamics of fixations depends on the image content. Also, various quantitative metrics have been developed in vision sciences to quantify differences and similarities between two eye tracking sequences.[84] Where required, these techniques can assist in understanding differences and similarities between radiological visual search patterns. For example, they can be used to identify pairs of radiologists who searched

the same mammograms in the most different fashion, with the aim of finding optimal pairings of radiologists.[51]

CONCLUSION

Given that breast cancer is the most common female cancer, more than 1 billion mammography examinations are annually performed and interpreted worldwide. Each year, radiologists also read a large volume of breast ultrasound, DBT, and breast MRI. Studies show extensive inter radiologist variability in their accuracy while reading breast images, and suggest that readers' characteristics, especially their experience and annual working load may influence their diagnostic accuracy.[85] Therefore, extensive visual search studies have been conducted to characterize how cognitive and perceptual processes of interpreting breast images relate to readers' characteristics. Evidence of these studies can provide guidance for enhancing practice in breast radiology and improving the diagnostic accuracy.

Findings of visual search studies could also assist educators to enhance training in breast radiology. The holistic processing model suggests that educators should present large volumes of normal and abnormal cases to students so that they build a good awareness of normal/abnormal appearances. Moreover, recent machine learning classifiers based on features extracted from fixation data were able to successfully distinguish less-experienced readers from experts. This opens new opportunities for monitoring the development of expertise and developing individualized training programs with different pace for each trainee's needs. Given the fact that widely accepted models for explaining expertise development have been suggested, designing good theory-driven studies of specific interventions in radiological education could be a possible future line of research.

Another important area addressed in medical image perception studies is mapping out the underlying reasons for errors in breast radiology. Despite the technological advances, which led to producing high quality images, error rates in radiology are still significantly high, and improving diagnostic errors in breast radiology is required. For example, the error rate in detecting cancer on mammograms varies from 10% to 30%, while approximately 80% of females recalled for additional imaging are ultimately diagnosed as normal.[85] Considering this high error rate, proper understanding of error causes could lead to error mitigation through developing appropriate educational interventions, a strategy for auditing and providing feedback to readers (*e.g.* computerized feedback system based on their visual search pattern). These solutions will potentially benefit millions of females each year. In summary, studies suggested that radiologists fixated on the majority of missed cancers but did not report them due to factors such as improper background sampling, incorrect first impression about the image, perceptual suppression of recognition, and prolonged visual search before making their decisions. In addition to the characteristics of the fixated area, the pattern of visual search itself relates to the diagnostic errors. Quantitative features extracted from mammographic scanpaths, describing their complexity, timing, and spatiotemporal characteristics were shown to be predictors of diagnostic errors. A few recent studies designed tools based

on eye tracking data and image content for detecting diagnostic errors. A potential direction for future research could be further improving these tools to make them compatible with clinically realistic viewing configurations. Current tools rely on features extracted from scanpaths recorded during the assessment of a single mammographic view or CC-MLO views. However, these experimental set-ups do not resemble the real practice, where four breast views, prior examinations, and different options for image manipulation and configurations for image presentation are available to the radiologists. Another important future line of research is exploring optimal ways to integrate the responses from such tools as well as CAD into radiology workflow.

A limited number of studies have been conducted to explore radiologists' visual search in imaging modalities such as breast ultrasound, DBT, and breast MRI. Considering the unique characteristics of visual search in each search task, future studies focusing on these modalities are highly recommended. For example, radiologists' scanpaths while reading DBT images cannot be fully categorized as either drillers' or scanners' search patterns, which were previously identified while searching for lung nodules in volumetric lung CTs. Moreover, it is unknown whether the holistic processing model proposed to describe radiological eye movements in 2D images can explain the radiologists' visual search patterns while examining a stack of images. Also, considering new findings emphasizing the importance of global image statistics in establishing the radiologists' first impression about the image, it is necessary to investigate how the holistic processing search model should be refined. Moreover, determining the intersection between models rooted in basic vision science and in medical image perception should be further investigated. Experimental and cognitive psychology can also offer different techniques to design simplified experiments for understanding the roles of attentional mechanisms and working memory. These techniques may be applied to design simplified experiments with non-radiologist participants which may possibly explain causes of diagnostic errors, role of attentional mechanisms, and better modelling of the medical image interpretation process.

COMPETING INTERESTS
None to declare.

REFERENCES

1. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? *Insights Imaging* 2017; **8**(no. 1): 171–82. doi: https://doi.org/10.1007/s13244-016-0534-1

2. Nakashima R, Kobayashi K, Maeda E, Yoshikawa T, Yokosawa K. Visual search of experts in medical image reading: the effect of training, target prevalence, and expert knowledge. *Front Psychol* 2013; **4**: 166vol.p.. doi: https://doi.org/10.3389/fpsyg.2013.00166

3. Krupinski EA. Current perspectives in medical image perception. *Atten Percept Psychophys* 2010; **72**(no. 5): 1205–17. doi: https://doi.org/10.3758/APP.72.5.1205

4. Sree SV, Ng EY-K, Acharya RU, Faust O. Breast imaging: a survey. *World J Clin Oncol* 2011; **2**(no. 4): 171vol.p.. doi: https://doi.org/10.5306/wjco.v2.i4.171

5. Drusch G, Bastien J, Paris S. "Analysing eye-tracking data: From scanpaths and heatmaps to the dynamic visualisation of areas of interest," *Advances in Science. Technology, Higher Education and Society in the Conceptual Age: STHESCA* 2014; **20**: 205vol.p..

6. Gandomkar Z, Tay K, Ryder W, Brennan PC, Mello-Thoms C. "Predicting radiologists' true and false positive decisions in reading mammograms by using gaze parameters and image-based features,". In: *Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment.*

9787: International Society for Optics and Photonics; 2016. pp. 978715.

7. Gandomkar Z, Tay K, Ryder W, Brennan PC, Mello-Thoms C. iCAP: an individualized model combining gaze parameters and image-based features to predict radiologists' decisions while reading mammograms. *IEEE Trans Med Imaging* 2017; **36**(no. 5): 1066–75. doi: https://doi.org/10.1109/TMI.2016.2645881

8. Tourassi G, Voisin S, Paquit V, Krupinski E. Investigating the link between radiologists' gaze, diagnostic decision, and image content," *Journal of the American Medical Informatics Association*. 2013; **20**(no. 6): 1067–75.

9. Voisin S, Pinto F, Morin-Ducote G, Hudson KB, Tourassi GD. Predicting diagnostic error in radiology via eye-tracking and image analytics: preliminary investigation in mammography. *Med Phys* 2013; **40**(no. 10): 101906vol.. doi: https://doi.org/10.1118/1.4820536

10. Carmody DP, Kundel HL, Toto LC. Comparison scans while reading chest images. taught, but not practiced. *Invest Radiol* 1984; **19**(no. 5): 462–6. doi: https://doi.org/10.1097/00004424-198409000-00023

11. Drew T, Evans K, Võ ML-H, Jacobson FL, Wolfe JM. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics* 2013; **33**(no. 1): 263–74. doi: https://doi.org/10.1148/rg.331125023

12. Swensson RG. A two-stage detection model applied to skilled visual search by radiologists. *Percept Psychophys* 1980; **27**(no. 1): 11–16. doi: https://doi.org/10.3758/BF03199899

13. Nodine C, Kundel H. "The cognitive side of visual search in radiology,". In: *Eye Movements from Physiology to Cognition: Elsevier*; 1987. pp. 573–82.

14. Kundel HL, Nodine CF, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology* 2007; **242**(no. 2): 396–402. doi: https://doi.org/10.1148/radiol.2422051997

15. Kundel HL, Nodine CF, Thickman D, Toto L. Searching for lung nodules. A comparison of human performance with random and systematic scanning models. *Invest Radiol* 1987; **22**(no. 5): 417–22. doi: https://doi.org/10.1097/00004424-198705000-00010

16. Krupinski EA. Visual scanning patterns of radiologists searching mammograms. *Acad Radiol* 1996; **3**(no. 2): 137–44. doi: https://doi.org/10.1016/S1076-6332(05)80381-2

17. Kundel HL, Nodine CF, Krupinski EA, Mello-Thoms C. Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Acad Radiol* 2008; **15**(no. 7): 881–6. doi: https://doi.org/10.1016/j.acra.2008.01.023

18. Sheridan H, Reingold EM. The holistic processing account of visual expertise in

medical image perception: a review. *Front Psychol* 2017; **8**: 1620. doi: https://doi.org/10.3389/fpsyg.2017.01620

19. Wolfe JM, Võ MLH, Evans KK, Greene MR. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn Sci* 2011; **15**(no. 2): 77–84. doi: https://doi.org/10.1016/j.tics.2010.12.001

20. Evans KK, Georgian-Smith D, Tambouret R, Birdwell RL, Wolfe JM. The GIST of the abnormal: above-chance medical decision making in the blink of an eye. *Psychon Bull Rev* 2013; **20**(no. 6): 1170–5. doi: https://doi.org/10.3758/s13423-013-0459-3

21. Evans KK, Haygood TM, Cooper J, Culpan A-M, Wolfe JM. A half-second glimpse often LETS radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proc Natl Acad Sci U S A* 2016; **113**(no. 37): 10292–7. doi: https://doi.org/10.1073/pnas.1606187113

22. Brennan PC, Gandomkar Z, Ekpo EU, Tapia K, Trieu PD, Lewis SJ, et al. Radiologists can detect the 'gist' of breast cancer before any overt signs of cancer appear. *Sci Rep* 2018; **8**(no. 1): 8717vol.p.. doi: https://doi.org/10.1038/s41598-018-26100-5

23. Gandomkar Z et al. "Does the strength of the gist signal predict the difficulty of breast cancer detection in usual presentation and reporting mechanisms?,". In: *Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment*. **10952**: International Society for Optics and Photonics; 2019. pp. 1095203.

24. Gandomkar Z et al. "Detection of the abnormal gist in the prior mammograms even with no overt sign of breast cancer,". In: *14th International Workshop on Breast Imaging (IWBI 2018*. **10718**: International Society for Optics and Photonics; 2018. pp. 1071804.

25. Kundel HL. "How to minimize perceptual error and maximize expertise in medical imaging,". In: *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*. **6515**: International Society for Optics and Photonics; 2007. pp. 651508.

26. Nodine CF, Kundel HL, Lauver SC, Toto LC. Nature of expertise in searching mammograms for breast masses. *Acad Radiol* 1996; **3**(no. 12): 1000–6. doi: https://doi.org/10.1016/S1076-6332(96)80032-8

27. Voisin S, Yoon H-J, Tourassi G, Morin-Ducote G, Hudson K. "Personalized modeling of human gaze: Exploratory investigation on mammogram readings,". In: *Biomedical Sciences and Engineering Conference (BSEC*; 2013. pp. 1–4.

28. Krupinski EA. "Medical image perception: evaluating the role of experience,". In: *Human Vision and Electronic Imaging V*. **3959**: International Society for Optics and Photonics.; 2000. pp. 281–90.

29. Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, et al. How experience and training influence mammography expertise. *Acad Radiol* 1999; **6**(no. 10): 575–85. doi: https://doi.org/10.1016/S1076-6332(99)80252-9

30. Gandomkar Z, Tay K, Brennan PC, Mello-Thoms C. "A model based on temporal dynamics of fixations for distinguishing expert radiologists' scanpaths,". In: *Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment*. **10136**: International Society for Optics and Photonics; 2017. pp. 1013606.

31. Gandomkar Z, Tay K, Brennan PC, Mello-Thoms C. Recurrence quantification analysis of radiologists' scanpaths when interpreting mammograms. *Med Phys* 2018; **45**: 3052–62. doi: https://doi.org/10.1002/mp.12935

32. Alamudun F, Yoon H-J, Hudson KB, Morin-Ducote G, Hammond T, Tourassi GD. Fractal analysis of visual search activity for mass detection during mammographic screening. *Med Phys* 2017; **44**(no. 3): 832–46. doi: https://doi.org/10.1002/mp.12100

33. Krupinski EA. "Influence of experience on scanning strategies in mammography,". In: *Medical Imaging 1996: Image Perception*. **2712**: International Society for Optics and Photonics.; 1996. pp. 95–102.

34. Mello-Thoms C, Nodine CF, Kundel HL. "Relating image-based features to mammogram interpretation,". In: *Medical Imaging 2002: Image Perception, Observer Performance, and Technology Assessment*. **4686**: International Society for Optics and Photonics.; 2002. pp. 80–4.

35. Mello-Thoms C, Dunn SM, Nodine CF, Kundel HL. The perception of breast cancers--a spatial frequency analysis of what differentiates missed from reported cancers. *IEEE Trans Med Imaging* 2003; **22**(no. 10): 1297–306. doi: https://doi.org/10.1109/TMI.2003.817784

36. Nodine CF, Mello-Thoms C, Weinstein SP, Kundel HL, Conant EF, Heller-Savoy RE, et al. Blinded review of retrospectively visible unreported breast cancers: an Eye-Position analysis. *Radiology* 2001; **221**, , (no. 1): 122–9. doi: https://doi.org/10.1148/radiol.2211001507

37. Mello-Thoms C. The problem of image interpretation in mammography: effects of lesion conspicuity on the visual search strategy of radiologists," *The British Journal of Radiology,* vol. 79, no. special_issue_2. 2006;: S111–S116.

38. Krupinski EA. "Visual Search of Mammographic Images: Influence of Lesion Subtlety1,". *Academic Radiology* **12**(no. 8): 965-969, 2005/08/01/ 2005..

39. Mello-Thoms C. The holistic grail: possible implications of an initial mistake in the reading of digital mammograms. In: *Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment*. **vol. 7263**: International Society for Optics and Photonics.; 2009. pp. 72630L.

40. Mello-Thoms C. How does the perception of a lesion influence visual search strategy in mammogram reading? *Acad Radiol* 2006; **13**(no. 3): 275–88. doi: https://doi.org/10.1016/j.acra.2005.11.034

41. Kundel HL. "Reader error, object recognition, and visual search,". In: *Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment*. **5372**: International Society for Optics and Photonics; 2004. pp. 1–12.

42. Nodine CF, Mello-Thoms C, Kundel HL, Weinstein SP. Time course of perception and decision making during mammographic interpretation," *American journal of roentgenology*. 2002; **179**(no. 4): 917–23.

43. Berbaum KS, Franken EA, Dorfman DD, Rooholamini SA, Kathol MH, Barloon TJ, et al. Satisfaction of search in diagnostic radiology. *Invest Radiol* 1990; **25**(no. 2): 133–9. doi: https://doi.org/10.1097/00004424-199002000-00006

44. Berbaum KS, Franken EA, Dorfman DD, Rooholamini SA, Coffman CE, Cornell SH, et al. Time course of satisfaction of search. *Invest Radiol* 1991; **26**(no. 7): 640–8. doi: https://doi.org/10.1097/00004424-199107000-00003

45. Samuel S, Kundel HL, Nodine CF, Toto LC. Mechanism of satisfaction of search: eye position recordings in the reading of chest radiographs. *Radiology* 1995; **194**(no. 3): 895–902. doi: https://doi.org/10.1148/radiology.194.3.7862998

46. Mello-Thoms C, Trieu PD, Brennan PC. "Going on with false beliefs: What if satisfaction of search was really suppression of recognition?,". In: *Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment*. **9037**: International Society for Optics and Photonics.; 2014. pp. 903708.

47. Mello-Thoms C, Hardesty L, Sumkin J, Ganott M, Hakim C, Britton C, et al. Effects of lesion conspicuity on visual search in mammogram reading. *Acad Radiol* 2005; **12**(no. 7): 830–402005/07/01/ 2005.. doi: https://doi.org/10.1016/j.acra.2005.03.068

48. Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP. "Role of computer-assisted visual search in mammographic interpretation,". In: *Medical Imaging 2001: Image Perception and Performance*. **4324**: International Society for Optics and Photonics.; 2001. pp. 52–6.

49. Selvi R. "Step-by-Step Approach to Read a Mammogram,". In: *Breast Diseases: Springer*; 2015. pp. 45–60.

50. Gandomkar Z, Brennan PC, Mello-Thoms C. "A cognitive approach to determine the benefits of pairing radiologists in mammogram reading,". In: *Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment*. **10577**: International Society for Optics and Photonics.; 2018. pp. 1057704.

51. Gandomkar Z, Tay K, Brennan PC, Kozuch E, Mello-Thoms C. Can eye-tracking metrics be used to better pair radiologists in a mammogram reading task? *Med Phys* 2018; **45**(no. 11): 4844–56. doi: https://doi.org/10.1002/mp.13161

52. Hamidinekoo A, Denton E, Rampun A, Honnor K, Zwiggelaar R. Deep learning in mammography and breast histology, an overview and future trends. *Med Image Anal* 2018; **47**: 45–67. doi: https://doi.org/10.1016/j.media.2018.03.006

53. Rodriguez-Ruiz A, Mordang J-J, Karssemeijer N, Sechopoulos I, Mann RM. "Can radiologists improve their breast cancer detection in mammography when using a deep learning based computer system as decision support?,". In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. **10718**; 2018.

54. Henriksen EL, Carlsen JF, Vejborg IM, Nielsen MB, Lauridsen CA. "The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review,". *Acta Radiologica* 2018; **0284185118770917**.

55. Gupta K, Sandhu P, Arora S, Bedi G. Role of high resolution ultrasound complementary to digital mammography. *Ann Afr Med* 2018; **17**(no. 3): 117vol.p.. doi: https://doi.org/10.4103/aam.aam_36_17

56. Gibbs V, Harrison G, Edwards H. Independent reporting sonographers: Could other countries follow the UK's lead?," *Imaging & Therapy Practice*. 2017;: 25–9.

57. Carrigan AJ, Brennan PC, Pietrzyk M, Clarke J, Chekaluk E. A 'snapshot' of the visual search behaviours of medical sonographers," *Australasian journal of ultrasound in medicine*. 2015; **18**(no. 2): 70–7.

58. Crofts G. A framework for guiding learning in ultrasound scanning. *Ultrasound* 2015;

23(no. 1): 6–10. doi: https://doi.org/10.1177/1742271X14562228

59. Kuhl CKet al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer," *Journal of clinical oncology*. 2005; **23**(no. 33): 8469–76.

60. Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-cancer screening--viewpoint of the IARC Working Group. *N Engl J Med* 2015; **372**(no. 24): 2353–8. doi: https://doi.org/10.1056/NEJMsr1504363

61. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. *JAMA* 1995; **273**(no. 2): 149–54.

62. Hooley RJ, Durand MA, Philpotts LE. Advances in digital breast tomosynthesis," *American Journal of Roentgenology*. 2017; **208**(no. 2): 256–66.

63. Aizenman A, Drew T, Ehinger KA, Georgian-Smith D, Wolfe JM. Comparing search patterns in digital breast tomosynthesis and full-field digital mammography: an eye tracking study," *Journal of Medical Imaging*. 2017; **4**(no. 4): 045501vol.p..

64. Dong L, Tang Q, Gale A, Bernardi D, Chen Y. "Analysis of visual search behaviour from experienced radiologists interpreting digital breast tomosynthesis (DBT) images: a pilot study,". In: *Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment*. **10577**: International Society for Optics and Photonics; 2018. pp. 1057707.

65. Timberg P, Lång K, Nyström M, Holmqvist K, Wagner P, Förnvik D, et al. Investigation of viewing procedures for interpretation of breast tomosynthesis image volumes: a detection-task study with eye tracking. *Eur Radiol* 2013; **23**(no. 4): 997–1005. doi: https://doi.org/10.1007/s00330-012-2675-z

66. Jiang Z, Das M, Gifford HC. Analyzing visual-search observers using eye-tracking data for digital breast tomosynthesis images. *Journal of the Optical Society of America A* 2017; **34**(no. 6): 838–452017/06/01 2017.. doi: https://doi.org/10.1364/JOSAA.34.000838

67. Drew T, Vo ML-H, Olwal A, Jacobson F, Seltzer SE, Wolfe JM. Scanners and drillers: characterizing expert visual search through volumetric images. *J Vis* 2013; **13**(no. 10): 3. doi: https://doi.org/10.1167/13.10.3

68. Venjakob AC, Mello-Thoms CR. Review of prospects and challenges of eye tracking in volumetric imaging," *Journal of Medical Imaging*. 2015; **3**(no. 1): 011002vol.p..

69. Kuhl CK, Schmutzler RK, Leutner CC, Kempe A, Wardelmann E, Hocke A, et al. Breast MR imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results. *Radiology* 2000; **215**(no. 1): 267–79. doi: https://doi.org/10.1148/radiology.215.1.r00ap01267

70. Rosen ELet al. Accuracy of MRI in the detection of residual breast cancer after neoadjuvant chemotherapy," *American Journal of Roentgenology*. 2003; **181**(no. 5): 1275–82.

71. Dao TH, Rahmouni A, Campana F, Laurent M, Asselain B, Fourquet A. Tumor recurrence versus fibrosis in the irradiated breast: differentiation with dynamic gadolinium-enhanced MR imaging. *Radiology* 1993; **187**(no. 3): 751–5. doi: https://doi.org/10.1148/radiology.187.3.8497625

72. Liberman L. *Breast MRI: diagnosis and intervention*: Springer Science & Business Media; 2005.

73. Wolfe JM, Horowitz TS, Kenner NM. Cognitive psychology: rare items often missed in visual searches. *Nature* 2005; **435**(no. 7041): 439vol.p.. doi: https://doi.org/10.1038/435439a

74. Wolfe JM, Horowitz TS, Van Wert MJ, Kenner NM, Place SS, Kibbi N. Low target prevalence is a stubborn source of errors in visual search tasks. *J Exp Psychol* 2007; **136**(no. 4): 623–38. doi: https://doi.org/10.1037/0096-3445.136.4.623

75. Consortium BCS. Cancer rate (per 1,000 examinations) and cancer detection rate (per 1,000 examinations) for 1,960,150 screening mammography examinations from 2002 to 2006 by age," ed. 2009;.

76. Mitroff SR, Biggs AT. The ultra-rare-item effect: visual search for exceedingly rare items is highly susceptible to error. *Psychol Sci* 2014; **25**(no. 1): 284–9. doi: https://doi.org/10.1177/0956797613504221

77. Fleck MS, Samei E, Mitroff SR. Generalized "satisfaction of search": adverse influences on dual-target search accuracy. *J Exp Psychol Appl* 2010; **16**(no. 1): 60vol.p.. doi: https://doi.org/10.1037/a0018629

78. Cain MS, Adamo SH, Mitroff SR. A taxonomy of errors in multiple-target visual search. *Vis cogn* 2013; **21**(no. 7): 899–921. doi: https://doi.org/10.1080/13506285.2013.843627

79. Cain MS, Mitroff SR. Memory for found targets interferes with subsequent performance in multiple-target visual search. *J Exp Psychol Hum Percept Perform* 2013; **39**(no. 5): 1398vol.p.. doi: https://doi.org/10.1037/a0030726

80. Dowd EW, Mitroff SR. Attentional guidance by working memory overrides salience cues in visual search. *J Exp Psychol Hum Percept Perform* 2013; **39**(no. 6): 1786vol.p.. doi: https://doi.org/10.1037/a0032548

81. Adamo SH, Cain MS, Mitroff SR. Satisfaction at last: evidence for the'satisfaction'account for multiple-target search errors. In: *Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment*. **vol. 10577**: International Society for Optics and Photonics; 2018. pp. 105770A.

82. Adamo SH, Ericson JM, Nah JC, Brem R, Mitroff SR. Mammography to tomosynthesis: examining the differences between two-dimensional and segmented-three-dimensional visual search," *Cognitive research: principles and implications*. 2018; **3**(no. 1): 17.

83. Anderson NC, Bischof WF, Laidlaw KEW, Risko EF, Kingstone A. Recurrence quantification analysis of eye movements. *Behav Res Methods* 2013; **45**(no. 3): 842–56.

doi: https://doi.org/10.3758/s13428-012-0299-5

84. Le Meur O, Baccino T. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behav Res Methods* 2013; **45**, , (no. 1): 251–66. doi: https://doi.org/10.3758/s13428-012-0226-9

85. Ekpo EU, Alakhras M, Brennan P. Errors in Mammography Cannot be Solved Through Technology Alone. *Asian Pac J Cancer Prev* 2018; **19**(no. 2): 291–301. doi: https://doi.org/10.22034/APJCP.2018.19.2.291