What do experts look at and what do experts find when reading mammograms?

Jeremy M. Wolfe,[a,b,*] **Chia-Chien Wu**,[a,b] **Jonathan Li**,[c] and **Sneha B. Suresh**[a]
[a]Brigham and Women's Hospital, Boston, Massachusetts, United States
[b]Harvard Medical School, Cambridge, Massachusetts, United States
[c]Melbourne Medical School, Melbourne, Victoria, Australia

Abstract

Purpose: Radiologists sometimes fail to report clearly visible, clinically significant findings. Eye tracking can provide insight into the causes of such errors.

Approach: We tracked eye movements of 17 radiologists, searching for masses in 80 mammograms (60 with masses).

Results: Errors were classified using the Kundel et al. (1978) taxonomy: search errors (target never fixated), recognition errors (fixated <500 ms), or decision errors (fixated >500 ms). Error proportions replicated Krupinski (1996): search 25%, recognition 25%, and decision 50%. Interestingly, we found few differences between experts and residents in accuracy or eye movement metrics. Error categorization depends on the definition of the useful field of view (UFOV) around fixation. We explored different UFOV definitions, based on targeting saccades and search saccades. Targeting saccades averaged slightly longer than search saccades. Of most interest, we found that the probability that the eyes would move to the target on the next saccade or even on one of the next three saccades was strikingly low (~33%, even when the eyes were <2 deg from the target). This makes it clear that observers do not fully process everything within a UFOV. Using a probabilistic UFOV, we find, unsurprisingly, that observers cover more of the image when no target is present than when it is found. Interestingly, we do not find evidence that observers cover too little of the image on trials when they miss the target.

Conclusions: These results indicate that many errors in mammography reflect failed deployment of attention; not failure to fixate clinically significant locations.

[DOI: 10.1117/1.JMI.8.4.045501]

Keywords: mammogram; useful field of view; eye movements; Attention; visual search.

Paper 20288RR received Oct. 28, 2020; accepted for publication Jun. 28, 2021; published online Jul. 13, 2021.

Suppose that you are looking for the word urgent in this page of text (other than in this sentence). How much of the scene would you look at in order to find your target? What if you did not find what you were looking for? Would you look at the "whole" scene to confirm that the target word was not there? In one sense, you have looked at the whole scene. The entire page is visible to you. We can take that as evidence that there is "non-selective" visual processing across the entire visual field[1] but that is not the right answer to the colloquial question of how much of the scene was "looked at." In a task such as an airport baggage screening, reading an x-ray, or, for that matter, proofreading this page, it would be obviously unacceptable to say that a brief glimpse of the visual stuff in an image constituted looking at the whole image. In a visual search task, finding the target or concluding that it is absent involves selecting portions of the image for processing beyond non-selective processing. The purpose of the present work is to use the socially important task of searching for breast cancer in mammograms to provide new insight into what it means to look at an image in this selective sense when conducting a search.

---

*Address all correspondence to Jeremy M. Wolfe, jwolfe@bwh.harvard.edu

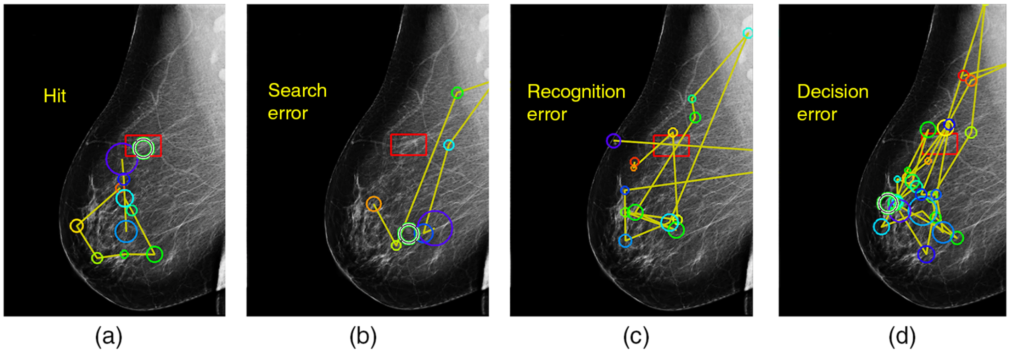


Fig. 1 Scan paths for four types of trials in a search for signs of breast cancer indicated by the red outline boxes: (a) A true positive ("hit") where the reader locates the target. Striped circle marks the reader's localization of the target. (b) A search error where the reader never fixates near the target. It is worth noting that in panel (b), the reader incorrectly marked a different location as a target, making this a miss error and a false positive error on the same case. (c) A recognition error where the reader fixates on the target for less than half a second and does not identify it correctly. (d) A decision error where the reader fixates on the target for more than half a second, but still does not identify it correctly.

The data for this study, described in more detail below, consist of eye tracking recorded from radiologists as they read a series of full-field digital mammograms of one breast. They were asked to "click" on a positive finding in each breast or to declare the breast to be normal. This produces a set of "scan paths",[2,3] examples of which are shown in Fig. 1.

Each colored circle in Fig. 1 shows one fixation when the eyes stopped at a location for some period of time. The duration of fixation is coded by the size of the dot. Connecting lines show the saccades, the ballistic movements of the eyes between fixations. Red outline boxes show the location of the lesion if present. Striped circles in Figs. 1(a), 1(b), and 1(d) show where the reader clicked to indicate the presence of a finding [albeit, incorrectly in Figs. 1(b) and 1(d)].

As noted, the goal of this work is to answer the colloquial question "what did the reader look at?" Some answers are obviously incorrect. Each of the 10 to 20 fixations in the examples in Fig. 1 shows the fovea being directed at one point. Clearly, we would not say that the reader looked at only 20 pixel-sized spots in the image. There is some region around each fixation that is looked at. That region can be called the "useful field of view" (UFOV) surrounding the point of fixation.[4,5] It could also be called the "functional visual field" (FVF).[6,7] Hulleman and Olivers[8] give a good standard definition as "the area of the visual field around fixation from which a signal can be expected to be detected given sensory and attentional constraints." They note that it is not an entity of fixed size. For instance, the area will be smaller if the signal to be detected is smaller and/or weaker. The terms UFOV and FVF are roughly equivalent. We will use UFOV as it is more common in the medical image perception literature. One of the purposes of this paper will be to show that the concept is more complex than we have typically thought and, in particular, that we can be misled when we "expect" a signal to be detected, just because it falls inside a plausibly defined UFOV.

Eye tracking allows us to propose richer accounts of what the reader looked at. Thus, in Fig. 1(a), the reader looked around, found the target, and clicked on it, without bothering to look at the upper right portions of the breast. It is worth noting that, in our task, the trial ended when an observer clicked on a finding, so we do not know if they would have looked more extensively if asked to find any and all signs of disease. In Fig. 1(b), the reader looked around and terminated the search without having found or fixated the target. In this specific example, the reader mistakenly responded to a non-target. In Fig. 1(d), not only did the reader decide not to look at some regions, they scrutinized the location of the finding (indicated by the red outline box in this image). In the end, they did not click on the target, making this a "decision error," to distinguish it from the "search error" in Fig. 1(b), where the target is never fixated. A third type of error, a "recognition error," is shown in Fig. 1(c). A recognition error is said to occur when the eyes land on or near the target but quickly leave again.[9]

In this study, we will use mammograms as our search stimuli. Mammography has been shown to be an effective tool for early diagnosis of breast cancer and can help reduce mortality.[10–12] However, while mammography has been the frontline screening tool for breast cancer for decades, there are high error rates, both false positives and false negatives. In the United States, ∼10% of screening mammograms are recommended for follow up for recall or biopsy (e.g., Refs 13–15). The use of digital breast tomography (DBT) may reduce recall rates,[14] but, given that the prevalence of disease is on the order 0.3% to 0.5% (e.g., Ref. 16), the vast majority of recalls can be considered false positive errors. In this case, calling these "false positives" does not mean that recalling these women was a mistake. Some features of normal parenchyma (breast tissue) look suspicious enough and ambiguous enough that further information is needed. Still, recalls result in significant anxiety, distress, and adverse impacts on the quality of life even once the woman learns that she does not have cancer,[17–19] so understanding if these can be reduced would be worthwhile.

False negative/miss errors can occur due to occult/hidden or subtle lesions where the reader could not be expected to see the lesion. Of more interest for our purposes are those false negatives that involve visible lesions that are, for whatever reason, not reported by the radiologist. These "retrospectively visible" lesions are defined as interval or screen-detected cancers that can be identified on the previous images once the location of the problem is known.[20,21] It is in these errors where the reader's search process is of most relevance. These errors are cancers that could have been detected but were not.

Efforts to reduce errors across different radiological search tasks can benefit from a characterization of the different types of errors made by the radiologist. Foundational work done by Kundel et al.[9] proposed that false negative errors in radiology could be placed into the three categories mentioned above (Fig. 1): search, recognition, and decision errors. The classification is based on the pattern of eye movements made by the radiologist. If the eyes never fixate on the lesion, it is classified as a search error. When the eyes do fixate on a target, the distinction between a "recognition" error and a "decision" error is based on how long the eyes dwell on the target. Kundel et al.[9] used 480 ms as the boundary. The original work, classifying errors in this manner, was done with search for lung nodules in chest x-rays and reported that 30% of errors were search errors, 25% were recognition errors, and 45% were decision-making errors. Krupinski[22] conducted a similar study of radiologists searching mammograms with similar results. The experts in her sample had 25% search, 25% recognition, and 50% decision errors. Her novice observers had 29% search, 42% recognition, and 29% decision errors. This study had a fairly small sample of errors (her three experts missed only eight cancers, in total, out of $3 \times 25 = 75$ lesions). An early study[23] reported that 43% of missed cancers were "overlooked." Presumably, these would be search errors.

Why are retrospectively visible cancers missed? There are two obvious factors: the cancer and the reader.[24] Unsurprisingly, more experienced readers typically do better,[25] reviewed in Ref. 26. Experienced observers learn where to look in much the same way that a new reader learns that there are places where it is important to look and places where it is not helpful. As a consequence, novices and experts tend to have different scan paths; differences that a computer classifier can detect.[27,28] That said, once the observers are trained, it appears that the case is more of a factor than the observer in determining what is missed. In one study, experienced readers were given a set of images with cancers that had or had not been missed in the clinic. The cancers that had been detected in the original screening, attracted attention more quickly than the retrospectively visible but originally missed cancers. This suggests that the missed cancers were, in fact, harder to find.[21] Beam et al.[29] directly compared case-related to reader-related factors and found that the case-related factors explained more of the disagreement between readers than did the reader-related factors; again, suggesting that harder cases are harder for everyone. Factors that may make detection and/or image interpretation difficult include the physical characteristics of the patient such as body habitus, high breast densities that may mask the cancer, breast augmentations/implants as well as lesion morphology, and location heterogenicity of the breast parenchyma.[23,30]

Observer expectations can also play an important role in medical image search as illustrated by an experiment by Drew et al. About 24 radiologists were asked to perform a lung nodule detection task in a CT image. Fully, 83% of radiologists did not report the presence of an image of a gorilla that had been embedded in the last trial in this experiment.[31] This gorilla was about 48 times larger than an average nodule and eye tracking revealed that the majority of radiologists who had missed the gorilla had fixated on or near it. The choice of a gorilla honored the gorilla featured in the most famous "inattentional blindness" study in the general psychology literature.[32] The result illustrates the point that "looking" at an item is not the same as registering, recognizing, and reporting it. Observers were looking for small, white round nodules and missed a big, black, shaggy gorilla, even when they fixated on it. Failures to detect incidental findings[33] may be related to inattentional blindness. In mammography, it is possible that a reader, currently focused on search for one target (e.g., a mass) might look at but not register another (e.g., calcifications).

We can now return to the topic of the UFOV. The eye tracker gives an estimate of the location of the point of fixation. It does not measure the region around that point that is processed during that fixation. Nor can it determine if the eyes are fixated on one object while attention is deployed to some other object. Under most circumstances, attention is probably deployed near the point of fixation, though that need not be the case—a fact that you can confirm by staring at this text while holding your arm out to your right and wiggling your finger. You can attend to that motion in the periphery even while your eyes are pointed elsewhere.

The wiggling finger is an extreme example. More typically, you will be attending to, and processing some region in the neighborhood of the current point of fixation. Estimating the size of that UFOV is important for medical image analysis research because without that estimate it is not possible to determine how much of an image was looked at by the radiologist. Classifying errors also requires some estimate of the area that was processed around the point of fixation. There have been a variety of previous efforts to estimate the UFOV, mostly in lung images. In a Kundel et al.[34] study of lung nodule detection in chest x-rays, the authors found nodules were detected well if fixations fell within 3.5 deg from the center of the nodule. Another study by Carmody et al.[35] also instructed readers to look for nodules within chest x-rays. The detection accuracy of lung nodules was reduced by one-half when the tumor nodules were located 5 deg from the center of the fixation. In the same paper, the authors identified a fixation duration of 300 ms as sufficient to detect 85% of the nodules when they were viewed directly, which may cast light on the distinction between recognition and decision errors. Some recent work has extended the study of the UFOV into three-dimensional (3D) volumes of image data in Lung CT[36] and DBT.[37] The 3D case changes the question from the percentage of the area of an image that has been examined to the percentage of the volume.

Interestingly, these studies of the UFOV in radiology have focused on the ability to identify targets at different positions in the periphery. This definition of the UFOV is based on the fall-off in acuity and the rise in crowding[38] as a function of eccentricity. This is a different question from the question of what was attended/examined during the search for that target. Inattentional blindness is clear evidence for this distinction. The gorilla could have been identified but it was not. It is possible for an item to be clearly resolvable yet not be noticed. We can make some progress in understanding what is processed during search by understanding that there are several different senses of UFOV that are in play at the same time in search. These UFOV variants are undoubtedly related (e.g., you are more likely to attend to an item in the periphery, if it can be recognized at that eccentricity.). However, the variants are logically distinct. As noted, most prior studies of the UFOV in medical image perception have been concerned with what can be called the resolution UFOV. If you are fixated at one point in an image, acuity, and crowding[38] considerations will make it possible to recognize some items but not others. Thus, in Fig. 2(a), if you are fixated on the X, you may be able to resolve the T at about 11 o'clock but not the more distant T at 8–9 o'clock, though it is of the same size. For that target type on this background, one could draw a contour representing the resolution UFOV. As Hulleman and Olivers[8] emphasize, the size of that UFOV would be different for different images. In an inhomogeneous stimulus such as a mammogram, the resolution UFOV will vary from location to location. When something like a 3.5 deg resolution UFOV is proposed for a task like lung nodule detection, it is necessarily some type of average over many locally different situations. If we look at Fig. 3(a), the lesion, marked by the red box, could be detected from a fixation point many degrees away. Of course, that does not mean that a smaller, lower contrast lesion in a dense, noisy breast could be detected at the same eccentricity.

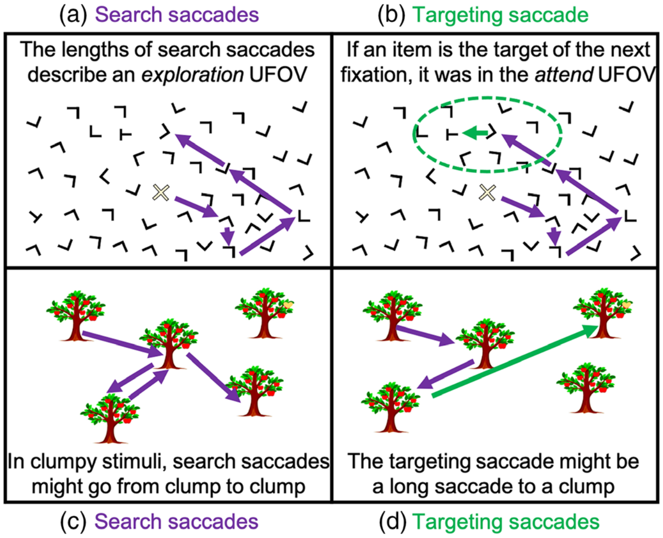


Fig. 2 Different types of saccade are related to different senses of the UFOV. Saccades that land on or near the target can be designated as “targeting saccades” (b) though some of those saccades might not involve detection of the target (see text for further discussion). Saccades that do not land near the target can be designated as search saccades (a). In a patchy display, like the orchard of (c) and (d), there may be larger saccades from tree to tree and smaller saccades within the tree/patch. The first saccade to the golden apple target (d) might be very similar to other long search saccades even if the target could not be resolved at that distance.

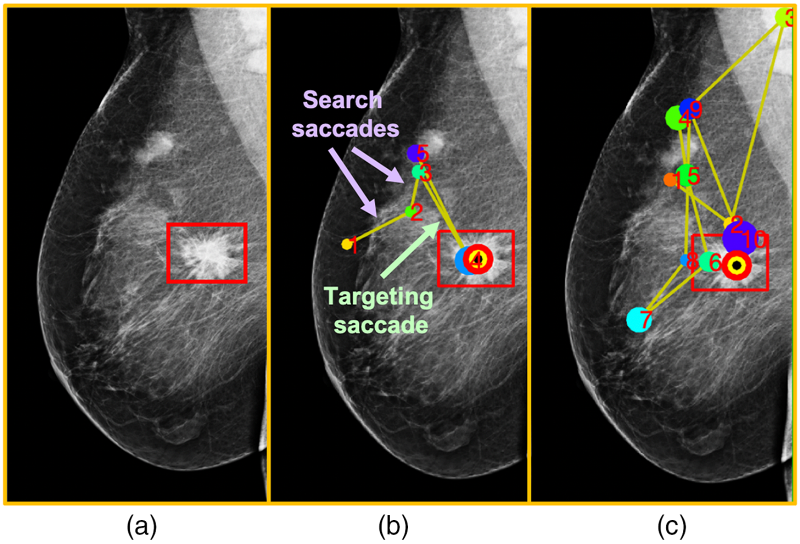


Fig. 3 Illustrative scanpaths for the mammogram shown in (a). (b) A fairly straight-forward scanpath where the reader makes some search saccades before fixating the target. (c) A more complex scan path, where the observer fixates the target on fixation #2, goes away, returning on fixation #6; goes away once more, returning a final time on fixation #10.

In addition to the resolution UFOV, we can also define an attentional UFOV and an exploratory UFOV. The attentional UFOV would be defined as the region around the current fixation where attention might be covertly deployed during the current fixation. The exploratory UFOV would be the region around the current fixation defined by the possible destinations of the next saccade. Across stimuli and search tasks, the three forms of UFOV will be correlated but are not the same. Thus, in Fig. 2(a), if one was fixated at the X, looking for a T, there are some letters in the vicinity of the X that are more likely to be attended than more remote Ts. It is likely, but not required that these deployments of covert attention will be constrained by the resolution UFOV. In Fig. 2(c), covert attention might visit a remote tree even if it were impossible to resolve a target apple in that tree. If one were fixated at the X in Fig. 2(a), the next fixation could go anywhere in the image (and many places beyond its borders). However, in a search task, it is likely that the next fixation will not land in some remote corner of the display. Instead, there will be a probability density function around the current fixation that defines the probable locations of that next fixation. As with the attentional UFOV, the exploratory UFOV need not be the same as the resolution UFOV. In Fig. 2(a), it would be possible to move the eyes to the T at 8–9 o'clock, even if that T were outside of the resolution UFOV. One last word about the attentional UFOV. We think of it as a region within which stimuli are entered into a processing pipeline in series.[39] Others would model this as limited capacity parallel processing throughout the entire attentional UFOV [e.g., Ref. 8].

It is important to stress that we are not proposing that resolution, attentional, and exploratory UFOVs should be thought of as different things such as lions, tigers, and bears. They are three logically distinct ways to think about what will happen next during a visual search when the eyes are fixated at one location in the image: What can be resolved, what can be attended, and where will the eyes go next? Ideally, we would like to be able to measure each of these types of UFOV. The resolution UFOV can be measured by standard psychophysical methods. Fixate at one location and measure the ability to detect/identify the target at locations around that point. Figure 2(b) offers a simple scheme for estimating attentional and exploratory UFOVs. Starting from X, the eyes make a set of saccades, ending on a target item. Consider the situation where the eyes are focused at the destination of the last of the purple/dark arrows in the sequence. The next saccade (green/light arrow) goes directly to a target. We can call that saccade a targeting saccade and refer to the purple saccades as search saccades. We could propose that the distribution of targeting saccades defines the attentional UFOV on the assumption that, in order to make a saccade to the target, that target must have been covertly attended while the eyes were at the prior fixation location. The distribution of the search saccades would define the exploratory UFOV.

We will show these distributions in the Results section; however, it must be noted that these distributions, while interesting, are imperfect tools for estimating the sizes of different UFOVs, especially the attentional UFOV. One problem is that there are, at least, three possible accounts of that green targeting saccade in Fig. 2(b). First, as proposed above, it could be that the target was detected while the eyes were fixated at the start point of that final saccade. Second, the eyes could be deployed to the next likely location of a target (a search saccade) and the observer would be pleasantly surprised to find that a target was fixated. This can be illustrated in the clumpy, orchard search of Fig 2(d). The green saccade goes to the target location, not because the target was detected but because that seemed to be a good place to look next. Even in an unclumpy display like Fig. 2(a), a letter could be selected simply because it could plausibly be a target, rather than because it had been covertly attended and identified. The final possibility is that the saccade could go to the target location by chance, though that is relatively unlikely.

As shown in Fig. 3, mammographic search is more like the orchards of Figs. 2(c) and 2(d) than the homogenous fields of Figs 2(a) and 2(b). Classic models of search in breast cancer screening [reviewed in Ref. 24] have emphasized the rapid assessment of the structure/gestalt of the image with fine-grained scrutiny coming only after a global assessment of the case. That gestalt impression could include rapid detection of the actual target, as in those lung nodules that could be detected in a 300 ms flash.[35] The gestalt would also include an assessment of where the trees were in this mammographic orchard and it might include a global assessment of how likely it would be that this mammogram would contain a lesion.[40]

Figure 3 shows the complications in interpreting targeting saccades in our data. Figure 3(a) shows a large finding. The red bounding box in the figure would not have been visible to the reader. Figure 3(b) shows a relatively uncomplicated scan path from one reader. The reader makes a couple of search saccades that would contribute to the estimate of the exploratory UFOV. Then the third saccade from fixation 3 to fixation 4 takes the eyes to the target. That targeting saccade would be considered part of the distribution defining the attentional UFOV. The account in Fig. 2(b) would assume that the reader recognized the target from the position of the third fixation. It is worth noting that there is an additional saccade, generated even after the target was correctly fixated. That final saccade away from the target would not be used to estimate either the exploratory or attentional UFOV. The situation in Fig. 3(c) is more complex. The first saccade goes to the target but the reader proceeds to make several saccades to the target before responding. Which saccade should be thought of as the true targeting saccade? We cannot know with certainty. Accordingly, we will separately analyze the first saccade to the target, the last saccade to the target, and the other saccades whose endpoints lie on or near the target. We will return to the complicated assessment of the attentional UFOV in the later sections.

The exploratory UFOV can be estimated in a more straight-forward manner from search saccades. Target absent/negative trials only contain search saccades so the distribution of those saccades can provide a reasonable estimate of the exploratory UFOV that can be compared with the distribution, on target present trials, of saccades prior to the first fixation on the target. To anticipate the results, those distributions are essentially the same.

1 Methods

Twenty-four readers were recruited at the 2019 Radiological Society of North America (RSNA) Conference and the data from seven participants were excluded due to poor or incomplete eye movement recordings. Of the remaining 17, six were radiologists specializing in mammography (or ‘experts’), seven were residents with some familiarity with mammography but limited experience. Two readers were radiologists with other specialties. One was a sonographer and we lack information about the 17th observer. Readers were tested in a darkened space in the Medical Image Perception Lab, built in the exhibit area of the RSNA meeting. Readers gave informed consent. They were asked to read 80 single-slice mammograms in the mediolateral oblique view. Of those 80 cases, 60 had a cancerous mass and 20 were normal. The high prevalence rate was designed to obtain sufficient samples of saccades to the target, recognizing that search strategies might be different if target prevalence was lower. Readers were asked to click on the lesion location or the “next” button if they thought the case was normal. Once the reader responded, the next case was presented without case-by-case feedback being given. Their eye movements were recorded during the experiment using a SMI RED 250 mobile eye tracker (SensoMotoric Instruments, Germany) with a refresh rate of 250 Hz. Observers conducted a nine-point calibration and did five practice trials before starting the experiment. Though no feedback was given after each trial, observers were able to review the collection of their false negative/miss trials at the end of the experiment.

1.1 *Images*

The stimuli were 80 full-field, left/right medio-lateral oblique digital breast mammograms obtained from the Dokuz Eylul Mammography Set,[41] which were used in Carrigan et al.[42] Of these, each of 60 cases had a single mass present. The other 20 were normal. All cases were previously diagnosed and given a score according to the Breast Imaging Reporting and Data System Atlas (BIRADS). The normal cases had a previously assigned BIRADS code of 1. The 60 abnormal cases with a single mass had BIRADS scores of 3 to 5. All stimuli were de-identified and presented on a 15-in eye tracker laptop screen. Observers sat ~50 cm away from the screen. The original resolution of the single mammogram was 3328×4096 pixels, which were then downsized to 878×1080 pixels (17.3×21.3 deg) in order to fit the eye tracker screen. The reduction of resolution might have reduced the visibility of very small features like calcifications, but these stimuli did not contain calcifications.

1.2 *Analysis*

Eye movements data were analysed into saccade and fixation using the SMI default event detection algorithm with saccade peak velocity of 40 deg/s and minimum fixation duration of 50 ms.

2 Results

Using this rich dataset, we will address four primary questions:

1. What are the distributions of the errors made by our observers?
2. What are the distributions of different types of saccades that can be used to define exploratory and attentional UFOVs?
3. What is “seen” or processed during a fixation?
4. How much of the image does a reader “look at” before making a decision about an image?

2.1 *Performance & Errors*

Basic performance statistics are shown in Fig. 4. As shown in Fig. 4(a), readers clicked on only about 60% of targets. This apparently poor performance is somewhat misleading. In this task, the trial ended when the observer clicked on any apparent lesion. Thus, there are two types of errors on target present trials. A reader could simply miss the target or they might click on some other item, creating a miss and false alarm (FA) error. In the case of a miss and FA, we cannot know if the reader would have gone on to find the true target as well. The experiment was simply not designed to allow multiple responses. On target absent trials, readers responded correctly on about 50% of trials, making false positive errors on the other trials. The percentage of false positive errors on absent trials is strongly correlated with the percentage of miss and FA errors on present trials ($r = 0.82$), suggesting that some observers were more likely to accept mass-like features in the display as masses. This high false positive rate is undoubtedly related to the high target prevalence rate in our stimuli.[43]

Response times (RT) are shown in Fig. 4(b). There is a clear effect of response type on RT (ANOVA: Greenhouse–Geiser corrected: $F(2.095, 33.52) = 11.19$, partial eta-sq = 0.41, $p = 0.0002$). Pairwise comparisons show that “hit” RTs are faster than each of the other types (all $p < 0.0004$). None of the other comparisons are significant. Thus, when the target is relatively easy to find, readers find it and are done quickly. The FA are apparently chosen more reluctantly or, at least, more slowly.

We find no effects of expertise in these data. We compared the 6 expert mammographers with the 7 residents and found no main effects of expertise on error rates (ANOVA: $\mathrm{F}(1, 43) = 0.1140$, partial eta-sq $= 0.00$ $p = 0.7373$) or RT ($\mathrm{F}(1, 55) = 0.9243$, partial eta-sq $= 0.02$, $p = 0.3405$) nor were the interactions of expertise with response type significant (both $p > 0.75$). This is a negative finding with not a great deal of statistical power. However, the lack of an effect could be seen as unexpected. It may be that even a trainee radiologist has mastered the basic task of detecting a mass in a mammogram. The effects of expertise might be seen more clearly when the reader needs to synthesize information from several sources.

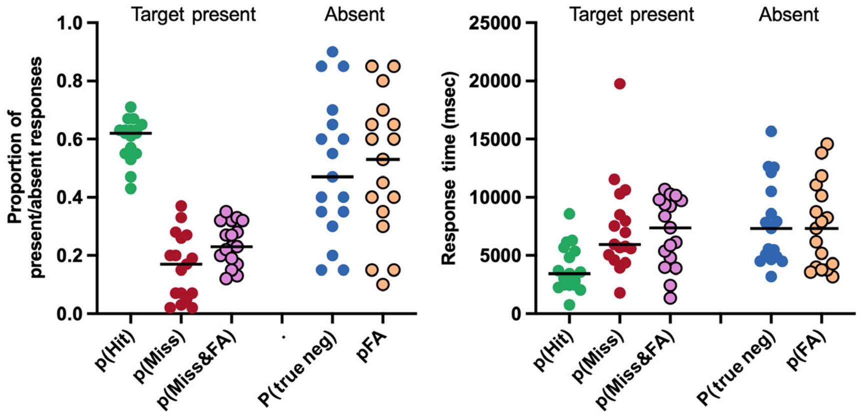


Fig. 4 (a) Accuracy measures: each datapoint represents one observer. Horizontal black lines show the mean. In this task, it is possible to miss a target or to miss a target and incorrectly identify a non-target as a target (miss and FA as shown in purple). (b) RT for each type of response.

Here, as in previous work,[29] performance appears to be driven by the images and not the observers. If we plot RT against accuracy for each target present case, we get a strong negative correlation ($r = -0.82$). Performance is fast and accurate for some images; slow and inaccurate for others. If we plot RT against accuracy for each observer rather than for each case, the correlation is minimal ($r = 0.08$). There is no evidence, for example, for fast, accurate experts in this task. Indeed, the residents are, on average, somewhat faster than the experts, though this is not statistically significant. This is not to say that expertise does not matter. It simply suggests that, in this straight-forward task, with these images, residents could perform about as well as experts. It could also be that the unfamiliar reading format and/or the lower-resolution images used in this task may have hindered experts' reading efficiency.

2.2 *Miss Error Classification*

One advantage of the low hit rate in this task is that there are a sizable number of miss errors to classify as search, recognition, or decision errors. A saccade was deemed to be a targeting saccade if it landed within the bounding box, around the target, or within 1.5 deg of the center of that box, whichever criterion was more liberal. In most cases, the 1.5 deg radius region provided the more liberal criterion. A miss error was classified as a search error if no saccades were landing in the bounding box or within 1.5 deg of the target center. If there were one or more saccades landing in or near the bounding box, we calculated the cumulative dwelltime on the target as the sum of those dweltimes. If that dwelltime was less than 480 ms as in Kundel et al.[9] we classified the error as a recognition error. If the cumulative dwelltime was longer, we classified the error as a decision error.

Results are shown in Fig. 5. There was an average of 34% search errors, 27% recognition errors, and 39% decision errors, a pattern of results quite similar to that of Krupinski.[22] Miss errors can be divided into pure miss errors and miss errors with an accompanying FA. Search errors accounted for 25% of the pure miss errors and 40% of the miss and FA errors. This is reasonable. If the reader was persuaded to click on a false target, the search could end before they ever happened to fixate on the real target. Decision errors accounted for 50% of the pure miss errors but only 31% of the miss and FA errors. Indeed, if we restrict analysis to the pure miss errors, the 25%-25%-50% division of errors is exactly the pattern produced by Krupinski's[22] experts.

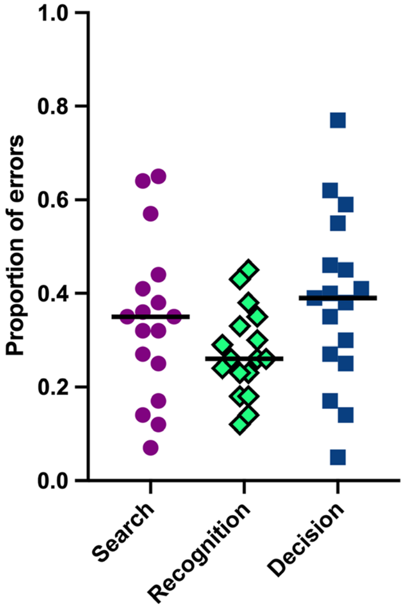
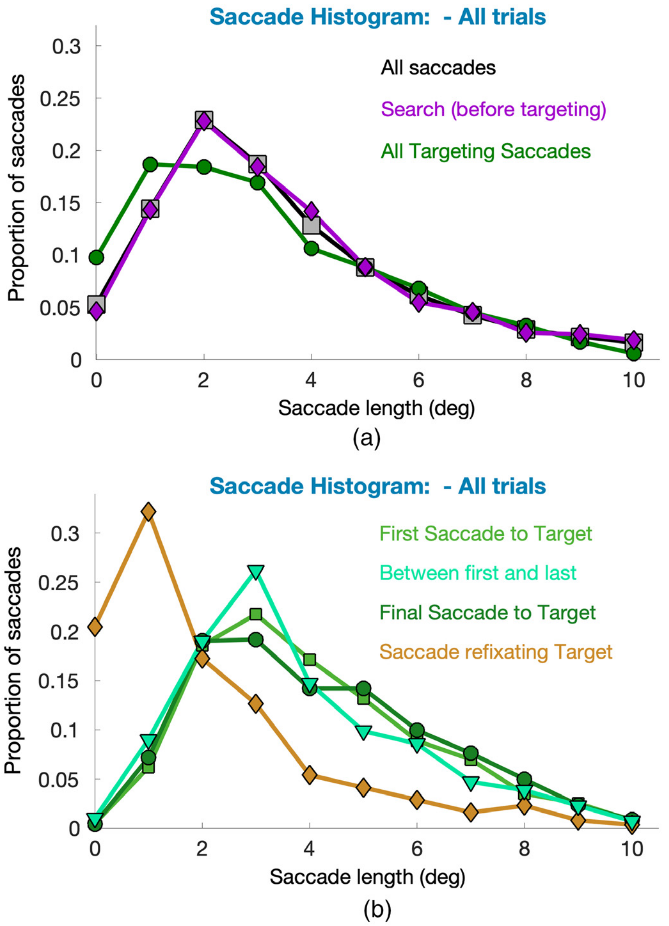


Fig. 5 Proportion of each of the three types of false negative error. Each data point shows results for one observer.

The relatively low level of performance in this task is not particularly important. It is a function of the images and of the method that allowed a single response per image. Our interest here is in the process of search and less in its outcome.

3 Saccade Amplitude Distributions and the Useful Field(s) of View

In a roughly 10 by 20 deg image as shown in the current experiment, readers could move their eyes to a randomly chosen spot 3 or 4 times a second as they would do in the natural viewing, but, of course, readers do not move randomly when reading a mammogram. Figure 6(a) shows a distribution with 1 deg-wide bins giving the distribution of all saccades in the experiment (black line). The plurality of saccades is in the 2 to 3 deg range. Because it is positively skewed, the distribution yields an average saccade length of 3.9 deg (s.d. $= 3.1$ deg, median $= 3.0$ deg, $N = 18{,}110$). Randomly distributed saccades over the same space would produce a positively skewed distribution that broadly peaks at between 5 to 8 deg with an average of 8 deg. Recall from Fig. 2, the exploratory UFOV will be defined by the search saccades, those saccades that do not go to the target. As we will describe, the attentional UFOV is more difficult to measure but is related to the saccades that go to the target. In Fig. 6(a), the green line shows the distribution of all saccades that go to the target. The purple line shows the distribution of all saccades that occur before the first saccade that lands on the target. These are "search saccades" and define the exploratory UFOV. This distribution is essentially identical to the distribution of all saccades



(a)

(b)

Fig. 6 Distribution of saccade length, pooled across observers. (a) All saccades (square), saccades before the first targeting saccade (diamond), and all saccades that end near the target (circle). (b) Subdividing targeting saccades: green lines (circle, square, and diamond) show three versions of saccades that come from some distance from the target to the target; brown diamonds show saccades that start on or near the target and end on or near the target.

(black) because observers are searching most of the time. All saccades in the negative/target absent cases could be considered search saccades though they are not included in the search saccade distribution in Fig. 6(a). For Fig. 6(a), there are 17,918 total saccades with saccade length equal to or smaller than 10 deg, 1885 targeting saccades, and 3870 search saccades (again, only those saccades that occur before the first saccade that lands on the target). Other saccades, such as saccades in the target absent cases and saccades occur after the targeting saccades, are included in the All Saccades distribution in Fig. 6(a).

Saccades that go to the target, in contrast, have a distribution that is significantly different from that of the search saccades [Kolmogorov–Smirnov statistic $(\text{KSstat}) = 0.75$, $p < 10^{-10}$]. Figure 6(a) shows what would seem to be a simple story. Observers make a series of search saccades. At some point, the target is near enough to the point of fixation that it attracts attention and the final, somewhat shorter saccade is directed to that target. That is what is shown in Fig. 2(b) and that is what is seen in eye movements recorded from observers searching for a T among Ls.[44] However, the story in this dataset is somewhat different and more complex. Figure 6(b) breaks the set of targeting saccades into four distributions as shown in the scan path in Fig. 3(c). There is a family of three green distributions that are all fairly similar. These are the distributions of the first saccade to the target, the last saccade to the target, and any extra saccades between the first and last. It is worth noting these are not mutually exclusive distributions since the first targeting saccade can also be the last. The first and last targeting saccades are actually longer on average than the search saccades (4.2 and 4.5 deg, respectively, for the first and last targeting saccades; 3.9 for the search saccades). The targeting saccades between first and last average 4.3 deg. All three of these targeting distributions differ from the search saccade distribution (all $\text{KSstat} > 0.12$, $p < 10^{-9}$) but not from each other (all $\text{KSstat} < 0.05$, all $p > 0.18$).

There is a fourth distribution of saccades with endpoints in the vicinity of the target and those are saccades that also start within the bounding box around the target or within 1.5 deg of the target center. Most of these are short saccades that we presume are refixations of the target as the observer attempts to figure out if the object of attention is, in fact, a target. If "refixating" saccades needed to start and end within 1.5 deg of the target center or inside the target bounding box, what is the source of the large saccades in the "saccade refixating target" distribution? Some saccades start on one side of the target and end up on the other side. Moreover, a few targets have bounding boxes that extend over 3 deg in one dimension. Thus, examining the edges of a large mass could lead to large saccades in this category. Only the refixating saccades have shorter saccade lengths on average than the search saccades. If those refixating saccades were removed from the targeting saccade distribution in Fig. 6(a), that green curve would shift slightly to the right of the purple search saccade distribution (Table 1).

The refixation distribution is clearly different from the distribution of search saccades and the other types of targeting saccades. That is of little interest since only shorter saccades are included in this distribution by definition. In addition, they might also be the by-product of the task as observers sometimes needed to look for the cursor after the lesion was fixated so they could click

Table 1 Lists all categories of saccades shown in Fig. 6.

Saccade category	Saccade subcategory
Search saccades: Saccades landing on non-target areas	Search saccades *before* the first targeting saccade (As shown in Fig. 6(a)
	Search saccades *after* the first targeting saccade (As shown in Fig. 7)
Targeting saccades: Saccades landing on the target (or within 1.5 deg of the target center)	First targeting saccade (could also be the last)
	Saccades between the first and last targeting saccades.
	Final targeting saccade (could also be the first)
	Saccade refixating the target

the lesion. In Fig. 6(b), there are 630 first saccades to the target, 683 last saccades to the target, 870 other targeting saccades, and 699 refixating saccades. The total number of targeting saccades is less than the numbers listed in Fig. 6(b) because, as noted, the categories of targeting saccades are not mutually exclusive.

Figure 7 shows the distributions of different classes of saccades that do not end at the target. Recall that Fig. 6(a) showed that the distribution of All Saccades is essentially identical to the distribution of search saccades before the target is found. Figure 7 explains that finding. Figure 7 replots the distribution for those search saccades in Fig. 6(a) (solid purple line). The dashed purple line shows the distribution of saccades between the first and last targeting saccade and the dashed dark red line shows the distribution of any non-targeting saccades after the first targeting saccade. In addition, the blue line shows all of the saccades on target absent trials, when, of course, there are no targeting saccades. It is clear that all of these are nearly identical to one another. There are small differences. Of the six comparisons between distributions, four are significant, with p-values between 0.0002 and 0.045. These are not corrected for multiple comparisons and, if they are significant, they are significant without being important.

The thin black lines in Fig. 7 are distributions of all saccades for individual readers. This gives a feeling for the relative consistency of the saccade distribution across individuals.

Figure 8 shows the spatial distribution of saccades by normalizing the endpoints of each saccade to 0,0 and tabulating the starting points of the saccades in 0.5×0.5 deg bins across a 10×10 deg field. Color indicates the fraction of saccades starting in that relative location with red indicating the highest density of starting points. The set of all bright, saturated points represents the most compact set of locations containing 75% of saccades. Dimmer, desaturated points represent the remaining 25% of saccade starting points. It is evident that all of these saccade types generate similar visualizations. If we take the envelope of the 75% points as an estimate, the fields are vertically oriented ellipses of about 12 deg vertically and 8 to 10 horizontally. One might have expected the ellipses to be horizontal as in typical visual field isopters [e.g., Ref. 45]. However, measured in this way, the fields appear to reflect the underlying vertical orientation of the mammogram. When we have used horizontally oriented stimulus arrays, we have obtained fields that are elongated in the horizontal direction.[44]

Figure 8(a) shows an estimate of the exploratory UFOV [see Fig. 2(a)]. Sadly, a similar estimation statement cannot be made for the attentional UFOV. The critical set of saccades would be those, cartooned in Fig. 2(b), where the target was found for the first time while the eyes were fixated at the starting point of the saccade. We cannot distinguish those targeting saccades from the saccades that landed on or near the target simply because that was the next interesting spot. As a concrete example, a 3 deg saccade to the target might show that the target was detected from 3 deg away and a subsequent 1 deg saccade that stayed near the target might be a refixation, made while the observer confirmed her decision or made her response. Alternatively, the same 3 deg saccade might have been a search saccade to a sensible spot and the subsequent 1 deg saccade might represent the moment of target recognition. It is simply not possible to know.

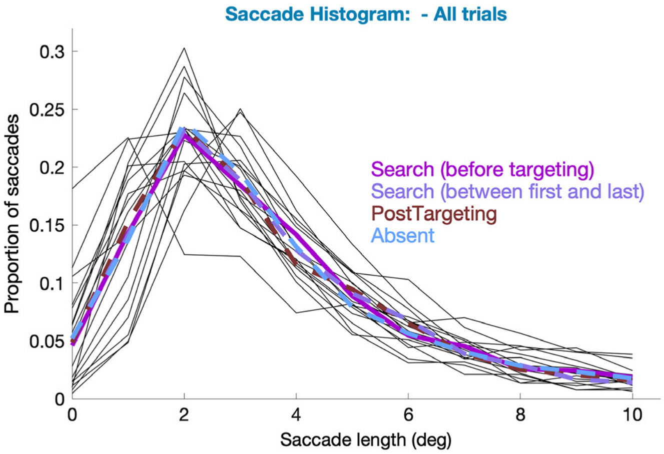


Fig. 7 Saccades that do not go to the target. All the distributions are similar, if not quite identical.

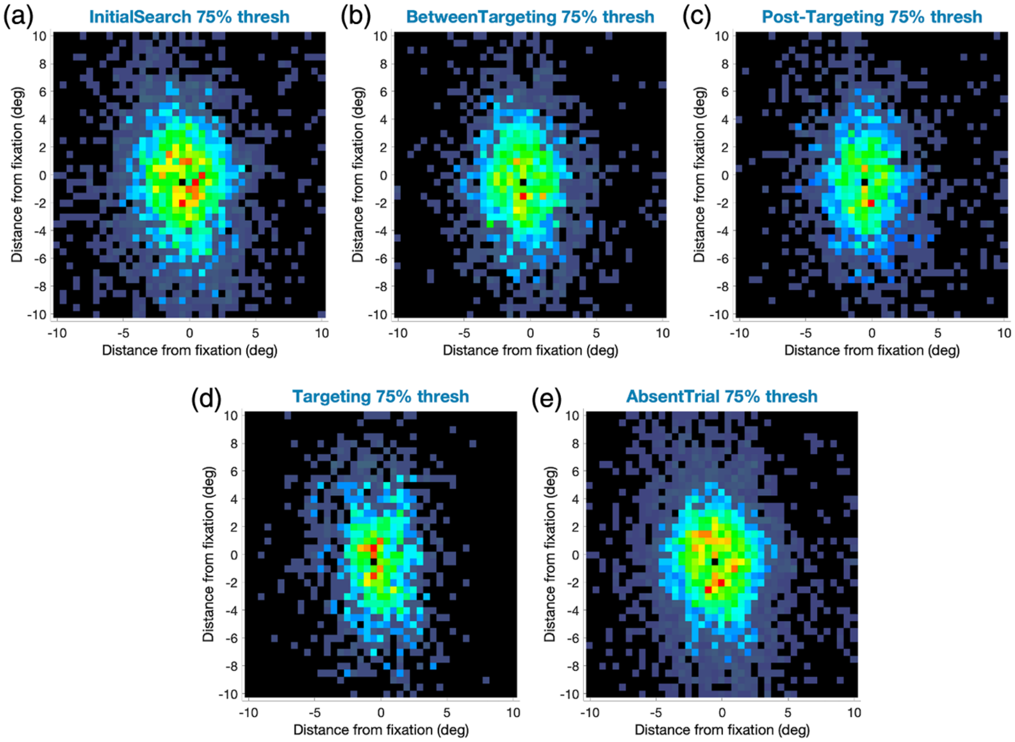


Fig. 8 Distributions of saccade starting points when ending points are aligned. (a) The search saccades defining the exploratory UFOV. (b) Saccades between the first targeting saccades and last targeting saccades. (c) Post-targeting saccades. (d) Targeting saccades. (e) Saccades in the target absent trials.

Thus, we can say that the attentional UFOV is contained in the distribution in Fig. 8(d) but we cannot more precisely define its shape on the basis of saccade length distributions. Looking at the overall targeting saccade distribution in Fig. 6(a), given that the average length of the targeting saccade is 3.4 deg and given that there were about 64% targeting saccades that had a length ≤ 3 deg, it would seem to be reasonably conservative to say that targets within 3 deg of fixation fall within an attentional UFOV.

3.1 *What Happens during a Fixation?*

Recall that Hulleman and Olivers[8] defined the UFOV (FVF) as "the area of the visual field around fixation from which a signal can be expected to be detected given sensory and attentional constraints." One could understand this to mean that during a fixation on one location, everything within the attentional UFOV around that location is processed. Based on the results of that processing, the target could be found, and targeted by a saccade or, if no target fell within the UFOV, the eyes would move to another location. However, as shown in Fig. 9, that is not the case for a large majority of saccades.

Figure 9(a) shows the probability that the next saccade will go to the ROI around the target, as a function of the distance from the current fixation to the target. For fixations prior to the first saccade to the target [These are the search saccades shown in Fig. 6(a)], the probability of the next saccade moving to the target is only about 30% even when the target is as close as 2 deg to the current fixation. Given the relatively large number of errors in this dataset, it could be that readers fail to fixate a nearby target on the trials when they simply fail to find the target at all. Perhaps they succeed in fixating the target on successful target present trials. The three curves in Fig. 9(a) show that this is not the case. To be sure, incorrect trials (red) produce lower rates of moving the eyes to the target but the effect is not dramatic and readers fail to move to the target on most saccades even on trials when they will, eventually, find the target.

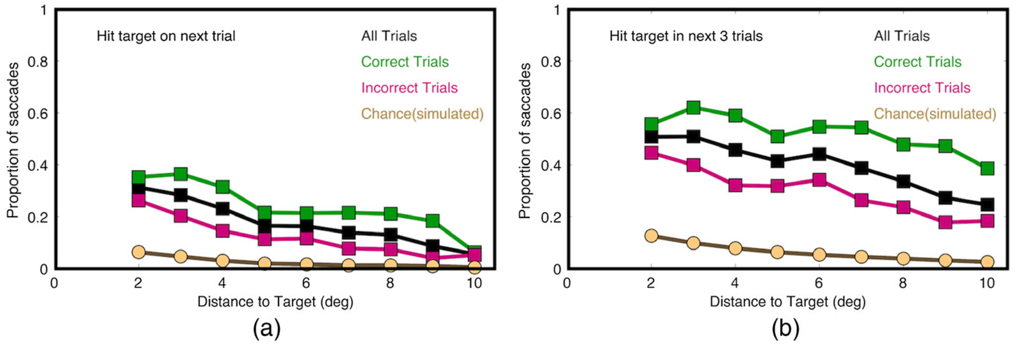


Fig. 9 (a) Probability that the next eye movement will go to the target as a function of the distance of the current fixation from the target. Green curve shows results for correct trials, black for all target present trials and red for incorrect trials. Beige shows a chance calculation described in the text. (b) Probability that at least one of the next three eye movements will go to the target as a function of the distance of the current fixation from the target.

Perhaps readers found the target when fixated nearby but took more than one subsequent saccade to fixate on that target. This might happen if the target was identified late in the current fixation, after the eyes were committed to a saccade elsewhere, for example.[46] Or the saccade to the target might have been inaccurate and required correction. To compensate for these possibilities, Fig. 9(b) shows the probability that any of the next three saccades land on the target. Obviously, this increases the chances of reaching the target but, even for correct trials, that probability rises to only about 0.6 for fixations 2 to 3 deg away from the target.

These results strongly suggest that the attentional UFOV around the point of fixation should not be considered to be a region within which everything is processed. Rather, it seems that the attentional UFOV defines an area within which information is sampled during the fixation. Returning to the metaphorical orchard of Figs. 2(c) and 2(d), when attention is directed to one apple tree, that does not mean that every apple from that tree is picked. It means that some of the apples can be picked before the observer moves on. In fact, we found a similar result in an experiment using T and L stimuli similar to those in the more uniform situation cartooned in Figs. 2(a) and 2(b). Within the attentional field surrounding the fixation, some letters will be sampled but, as in Fig. 9, we found that the probability that a nearby target will be fixated by the next saccade is well below 1.[44] The harder the task, the lower the probability. This result provides one possible explanation for the missed gorillas of Drew et al.[31] Even though fixation might have been nearby, a gorilla inside the UFOV is not necessarily a gorilla that is identified.

Perhaps we are substantially overestimating the size of the UFOV. Hulleman and Olivers[8] argue that the UFOV (they use the FVF terminology) shrinks as the task gets harder. With a hard task, the UFOV shrinks to the size of the fovea and only one item would be processed with each fixation. In the absence of any ability to recognize a target beyond the fovea, it would be no surprise if the next saccade did not reliably go to the target, even if the target were just two degrees away. If the eyes were, in fact, deployed randomly, the surprise would be that the percentage of saccades going to the target is at high as it is.

Even randomly deployed saccades would not be completely random, in the sense that they would not go to task irrelevant locations such as blank portions of the screen. What would performance look like if Os were making saccades to plausible locations but without having covertly identified a target? To estimate this, for every fixation in the dataset, we randomly chose a target location from the set of 60 target locations available from our cases. Using these locations roughly captures that fact that lesions are more likely in some locations than others. We measured the distance from the current fixation to the target and then determined if the next fixation [Fig. 9(a)] or any of the next three fixations [Fig 9(b)] landed on the randomly chosen target location. We eliminated saccades of <1.5 deg in length (already on the target) and >10 deg. The results are shown as the beige functions in Fig. 9. There is some probability of fixating the target by chance, but it is quite low. One can imagine other ways of computing chance performance, but we take this simulation as evidence that readers are not moving their eyes to random, if plausible locations, even if they are also not reliably moving their eyes to nearby targets. They appear to be imperfectly processing information in a UFOV and moving to the target if and when they find it.

3.2 *Estimating Coverage*

To estimate how much of the image was seen, we use the search saccade data from Fig. 9(a) (black squares). From this, we subtract the chance function from Fig. 9(a) (beige circles). This produces a probability of "seeing" function with a maximum value of about 0.25 at fixation,

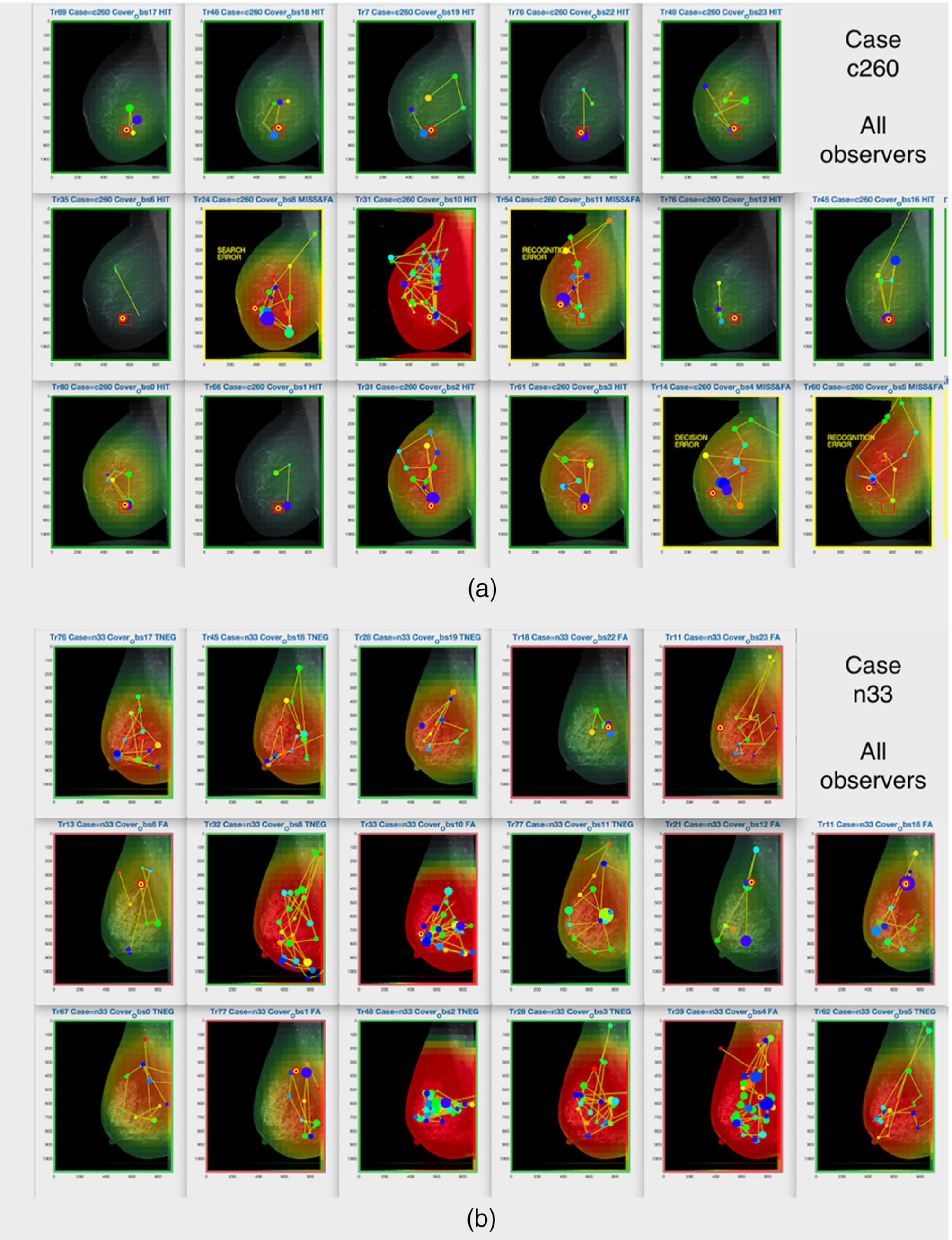


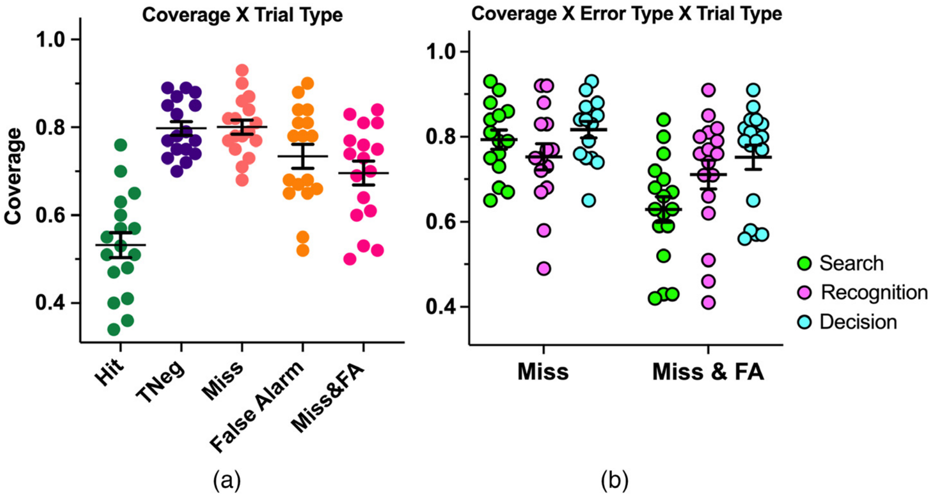
Fig. 10 Scan paths and coverage heatmaps for two representative cases for each of 17 observers: (a) target present, (b) target absent. Color of fixations gives order of fixation from red (first) through yellow and green to blue (last). Size of dots reflects dwelltime. Heatmap color reflects the probability that a region was seen for purposes of finding a target. Red ($p = 1.0$) to green to black ($p = 0.0$). Similar figures for all cases are available online.[47]

falling to 0.05 at ∼10 deg from fixation. We place this probabilistic UFOV at each fixation and calculate how likely it is that a location is processed whenever the UFOV covers that location. Thus, if one fixation produced a probability of 0.3 that information at some specific location would be processed and another fixation produced a probability of 0.2 for the same location, we would consider the probability of being seen at that location to be $1 - (1 - 0.3) * (1 - 0.2) = 0.44$. We repeat this process for all fixations and produce heatmaps of the results. It is worth noting that this assumes that deployments of covert attention are independent from fixation to fixation. This is probably an over-simplification. We could do the same analysis with the data in Fig. 9(b). This would simply increase the estimate of coverage. It does not change the relative coverage for different types of trials (as discussed below).

Figure 10 shows two representative cases: Fig. 10(a) shows target present (target location is indicated by the red outline box); Fig. 10(b) shows target absent. In Fig. 10(a), notice that there are actually a larger number of eye movements and more coverage on the error trials. As shown below, in this dataset, miss errors do not seem to be the product of a cursory search in which the reader quits the search too quickly. Notice that, as would be expected, coverage and number of fixations are greater for target absent cases Fig. 10(b). The effects of trial type and error classification are summarized in Fig. 11.

There is a main effect of trial type [ANOVA (d.f. Greenhouse–Geisser corrected): $F(2.674, 42.78) = 59.46$, partial eta-sq = 0.79, $p < .0001$] driven primarily by the smaller coverage for hit/true positive responses. Tukey's multiple comparison test shows the coverage in the "hit" condition to be smaller than each of the other conditions (all $q > 15$, all $p < 0.001$). Coverage is somewhat lower in cases where the observer makes an FA. The comparisons of the miss and FA cases to true negatives and simple "miss" trials are significant (both $q > 6$, $p <= 0.0025$). The comparisons of true negatives and simple miss trials to simple FA trend in the same direction ($p = 0.09$). This is probably a by-product of the method. If observers selected an incorrect location, that selection terminated the trial. Were observers permitted to search longer, they might have looked at more of the image before quitting.

Figure 11(b) shows coverage as a function of the error classification on miss and miss and FA trials. One might have expected the coverage to be lower for search errors since a search error is defined by a failure to get the eyes to within 1.5 deg of the target. However, that does not appear to be strongly supported in these data. There are significant main effects of trial type [$F(1, 90) = 15.18$, partial eta-sq = 0.15, $p = 0.0002$] and error type [$F(2, 90) = 3.523$, partial eta-sq = 0.07, $p = 0.0336$]. The interaction is weaker [$F(2, 90) = 2.648$, partial eta-sq = 0.06, $p = 0.0763$] that reflects the somewhat lower coverage for search errors on the miss and FA. Again, this may be a methodological issue since searches terminate after the FA response. Of most interest here is the finding that readers looked at the same amount of the image on plain miss trials, regardless of the classification of the error.



(a) (b)

Fig. 11 (a) Coverage as a function of trial type. (b) Coverage as a function of error classification for two versions of miss error.

This method of calculating coverage relies on the assumption that the coverage can be estimated using the probability that the next saccade will find the target [Fig. 9(a)]. It also makes use of an estimated guessing function. In order to determine if the patterns in our coverage results are robust, we performed the same analysis with two very different UFOVs. For one version, the UFOV had a 5 deg radius, based on earlier work.[35] Given our results, it seems implausible to imagine that observers would find any target that fell within that UFOV so we set a uniform probability of 0.7 of processing all locations inside this 5 deg radius UFOV. Some models do assume complete processing of the contents of the UFOV.[8] Thus, for the second UFOV, we used a UFOV radius of just 2 deg radius but assumed that everything within that radius was processed. The 5 deg UFOV produces results that are very similar to the coverage data presented in Figs. 10 and 11. The 2 deg produces markedly smaller coverage estimates, as would be expected. However, both of these alternate UFOVs produce the same relationship between coverage in different conditions with an essentially identical pattern of statistical results. Thus, hit/true positive trials have lower coverage than other trial types. Miss/false negative trials look like true negative trials. We see only weak evidence for miss errors that are due to observers failing to look at enough of the image. Even errors classified as search errors only show reduced coverage on trials that have both miss and FA errors. As noted, we suspect that coverage may be artificially reduced because the trial ended with the click on the false target.

Figure 12 shows the limitations of the use of a UFOV-based estimate of what the observer has looked at. The figure shows six search errors, defined by the failure of the eyes to land within 1.5 deg of the target (shown in red outline). The top row shows what we might call "sensible" search errors. Based on the analysis described above, the probability that the target would be seen was comfortably <1 and the target was, in fact, missed. The bottom row shows search errors where the probability of seeing the target is calculated to be near 1. Still, the observer missed the target. Looking at the scan paths, we can hypothesize that, even though the UFOV covered the

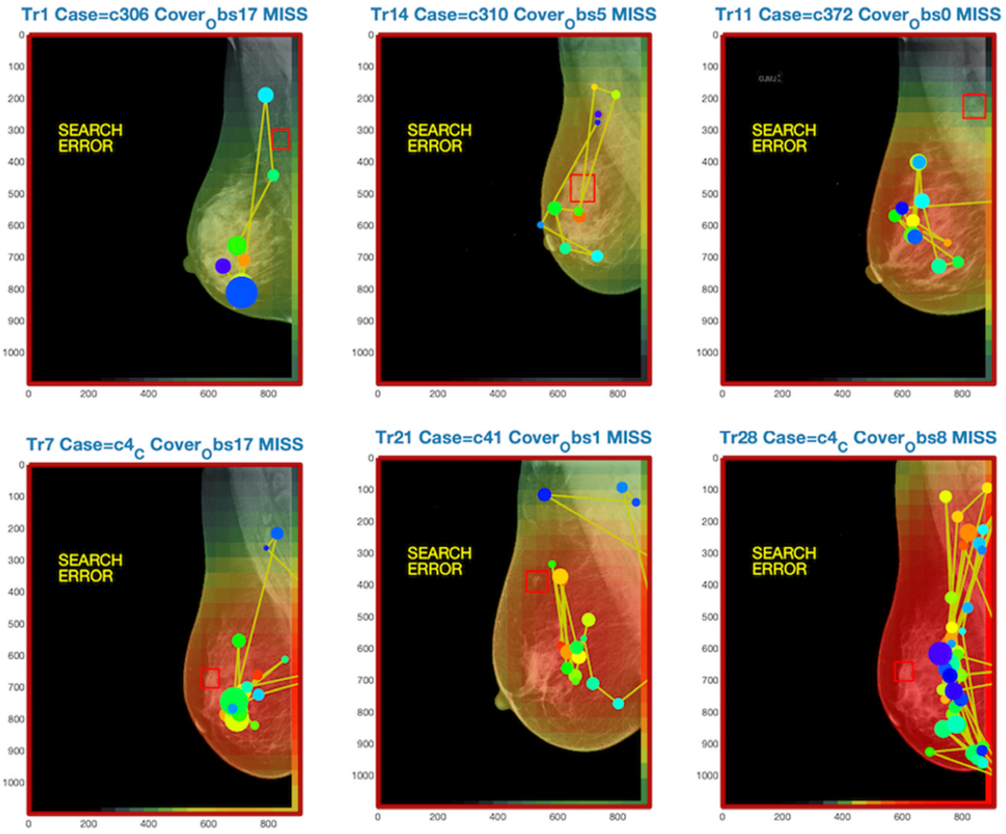


Fig. 12 Six search errors. The eyes are never closer than 1.5 deg to the target (red boxes). In the top row, UFOV seeing probability is markedly <1. In the bottom row, probability is ~1 but the item is still missed.

target, the observer's covert attention was directed elsewhere. It is probably an oversimplification to think of a round (or oval) UFOV where the probability of seeing is flat or declines smoothly from the point of fixation.

4 General Discussion

Returning to the four questions raised at the beginning of the results, we can offer some conclusions.

1. What is the distribution of the errors made by our observers?

Using the classic classification of errors in search, recognition, and decision, our observers produced a convincing replication of the classic results from Krupinski.[22] Critically, our coverage data did not provide strong evidence for a failure to look at enough of the image. Observers' coverage on miss trials looks like their coverage on true negative trials except when search is terminated by a false positive error and the reader misses a target on the same trial. It is possible that this effect of misses on false positive trials reflects a version of a satisfaction of search error, where finding one item makes it less likely that the observer will find another on the same trial[48–51] but the present study is not designed to test this hypothesis. It seems more likely that this effect is a side-effect of terminating the trial when the false target is selected.

2. What are the shapes and extents of the saccade-defined exploratory and attentional UFOVs?

On logical grounds, we can distinguish three types of UFOV that are relevant during a search task. These may not be separate things. They are better thought of as three ways to think about the UFOV. The resolution UFOV reflects the acuity and crowding limitations on what can be seen once attention is directed to a location. The exploratory UFOV describes the visual search strategy used to make overt deployments of the eyes. Finally, the attentional UFOV describes the area within which covert attention can be deployed while the eyes are fixated. In this study, the exploratory UFOV was mapped by the distribution of search saccades as shown in Figs. 6–8. The attentional UFOV is more difficult to map because, while finding a target is very likely to produce a saccade to that target in this task, the situation is asymmetrical. A saccade to the target does not guarantee that the target was covertly detected during the previous fixation. The saccade to the target could have been programmed merely as a saccade to a plausible location for a target. Figures 6 and 8 indicate that the saccade that represents the finding of the target is drawn from a distribution that is not radically different from the search saccades. This makes sense. The search saccades should be designed to get the eyes to a position where the target can be found.

3. What is seen or processed during a fixation?

Perhaps the most interesting aspect of these data is captured in Fig. 9. The figure shows the probability that the eyes will move to the target as a function of target eccentricity. That probability is strikingly low, even when the eyes are quite near to the target and even when we ask if any of the next three saccades go to the target. Clearly, it is too simple to propose that observers move their eyes around the image until the target falls inside the UFOV, at which point is it detected, fixated, and responded to. Instead, it appears that the eyes are used to forage in likely regions of the image, but, as in other foraging tasks (e.g., human berry picking or honey bees visiting a patch of flowers), arrival at a location does not imply the collection of every target near to that location.[52] Instead, a foraging animal[53] or human[54] will collect some of the available targets and then move on when the rate of return drops to some threshold.

It would be interesting but difficult to determine exactly what is processed on each fixation as an expert searches a mammogram. The coverage maps of Fig. 12 suggest that processing is not uniform or symmetrical around fixation. A more detailed map of covert attention would be hard to produce. As evidence for the difficulty of the task, consider the effects of dwell time. One might expect that the chance of moving to the target on the next saccade would increase as a function of the amount of time spent at the current point of fixation. However, when we divided the saccades by duration of the current fixation, we found that, if anything, longer dwell times made it less likely that the next saccade would go to the target.

Regardless of the precise details, the results shown in Fig. 9 shed light on dramatic phenomena such as "inattentional blindness." It is clearly possible to fixate near a target without processing it, even the target is at a location, relative to fixation, where it could be clearly recognized. The present results show that this is true, even if the missed item is the item that the observer is searching for. If we take the case of an incidental finding that is not the prime object of search (such as a gorilla in lung CT images, as in Drew et al.[31]), the situation is that much worse. Attention will be guided toward features of the target. The gorilla, with its incorrect features is even more likely to be missed, even when it falls inside an attentional UFOV.

4. How much of the image does a reader look at before making a decision about an image?

At the outset, we asked how much of the image is seen by an expert. It should be clear that this turns out to be a more difficult question than one might have imagined. As shown in Figs. 10 and 11, UFOV-based estimates would say that observers look at most of the image when a target is not present. However, they also look at about the same fraction of the image when the target is present, but not found. We did not find evidence for observers quitting searches dramatically too early. Even search errors, those where the point of fixation never got closer than 1.5 deg to the target, do not seem to involve early quitting except when the reader found a different, albeit incorrect, target (the miss and FA trials) and that effect is probably a by-product of the method.

This result might change if we repeated the task at low prevalence.[55] Observers, including radiologists,[56] miss more targets when those targets are rare and there is evidence that observers tend to quit more quickly when most cases are negative.[43] That said, forcing non-experts to spend more time on low-prevalence searches did not improve performance.[57] Overall, the key to understanding why people miss targets may lie less in what they do with their eyes but more in the more mysterious question of what they do with their attention. These results make it clear that getting the eyes to the right spot may be necessary, but it is not sufficient.

4.1 *Limitations*

As an effort to understand how radiologists might fail to report findings that turn out to be retrospectively visible, this study has some important limitations. It departs in many ways from normal clinical practice. Observers are looking at just one breast image without access to priors or other views. They know that the prevalence is much higher than it would be in a breast cancer screening setting. They also know that they are looking only for the presence or absence of a single mass. These and other differences might change the observer's search strategy. It would be valuable to collect a similar dataset with instructions to search for an unknown number of findings. This would allow us to assess search behavior after the reader finds the first target or after a false positive error. A richer understanding of search behavior would require that the reader could access other views of the breast, including priors. As eye tracking technology becomes ever more robust and less obtrusive, we can hope for a time when it will be practical to collect eye movement data of this sort from radiologists as they do their routine clinical work.

Disclosures

The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose

Acknowledgments

We thank Ann Carrigan for the use of the images used in this study. The study was conducted as part of the Medical Image Perception Lab, hosted at the annual meeting of the Radiological Society of North America (RSNA). We thank RSNA and Barco for in-kind support and NIH-NCI for financial support of the Perception lab. This research was supported by NIH-NCI grant CA207490.

References

1. J. M. Wolfe et al., "Visual search in scenes involves selective and non-selective pathways," *Trends Cognit. Sci.* **15**(2), 77–84 (2011).
2. D. Noton and L. Stark, "Scanpaths in saccadic eye movements while viewing and recognizing patterns," *Vision Res.* **11**(9), 929–942 (1971).
3. D. Noton and L. Stark, "Scanpaths in eye movements during pattern perception," *Science* **171**(3968), 308–311 (1971).
4. N. H. Mackworth, "Visual noise causes tunnel vision," *Psychonom. Sci.* **3**, 67–68 (1965).
5. R. Sekuler and K. Ball, "Visual localization: age and practice," *J. Opt. Soc. Am. A* **3**(6), 864–868 (1986).
6. A. F. Sanders, *The Selective Process in the Functional Visual Field*, Van Gorcum, Assen (1963).
7. A. F. Sanders, "Some aspects of the selective process in the functional visual field," *Ergonomics* **13**(1), 101–117 (1970).
8. J. Hulleman and C. N. L. Olivers, "The impending demise of the item in visual search," *Behav. Brain Sci.* **40**, e132 (2017).
9. H. L. Kundel, C. F. Nodine, and D. Carmody, "Visual scanning, pattern recognition and decision-making in pulmonary nodule detection," *Invest. Radiol.* **13**(3), 175–181 (1978).
10. B. Lauby-Secretan et al., "Breast-cancer screening–viewpoint of the IARC Working Group," *N. Engl. J. Med.* **372**(24), 2353–2358 (2015).
11. C. S. Lee et al., "Risk-based screening mammography for women aged <40: outcomes from the national mammography database," *J. Am. Coll. Radiol.* **17**(3), 368–376 (2020).
12. D. L. Monticciolo et al., "Breast cancer screening for average-risk women: recommendations from the ACR commission on breast imaging," *J. Am. Coll. Radiol.* **14**(9), 1137–1143 (2017).
13. S. Gottlieb, "Rate of recall after breast cancer screening in the United States is twice that in the United Kingdom," *Br. Med. J.* **327**(7421), 948–948 (2003).
14. E. L. Honig et al., "Factors impacting false positive recall in screening mammography," *Acad. Radiol.* **26**(11), 1505–1512 (2019).
15. B. L. Sprague et al., "Assessment of radiologist performance in breast cancer screening using digital breast tomosynthesis vs digital mammography," *JAMA Network Open* **3**(3), e201759 (2020).
16. C. S. Lee et al., "The national mammography database: preliminary data," *Am. J. Roentgenol.* **206**(4), 883–890 (2016).
17. C. Lerman et al., "Psychological and behavioral implications of abnormal mammograms," *Ann. Intern. Med.* **114**(8), 657–661 (1991).
18. I. T. Gram, E. Lund, and S.E. Slenker, "Quality of life following a false positive mammogram," *Br. J. Cancer* **62**(6), 1018–1022 (1990).
19. J. Brett, J. Austoker, and G. Ong, "Do women who undergo further investigation for breast screening suffer adverse psychological consequences? A multi-centre follow-up study comparing different breast screening result groups five months after their last breast screening appointment," *J. Public Health Med.* **20**(4), 396–403 (1998).
20. B. Boyer et al., "Retrospectively detectable carcinomas: review of the literature," *J. Radiol.* **85**(12), 2071–2078 (2004).
21. C. F. Nodine et al., "Blinded review of retrospectively visible unreported breast cancers: an eye-position analysis," *Radiology* **221**(1), 122–129 (2001).
22. E. A. Krupinski, "Visual scanning patterns of radiologists searching mammograms," *Acad. Radiol.* **3**(2), 137–144 (1996).
23. R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**(3), 613–617 (1992).
24. Z. Gandomkar and C. Mello-Thoms, "Visual search in breast imaging," *Br. J. Radiol.* **92**, 20190057 (2019).
25. C. F. Nodine et al., "Nature of expertise in searching mammograms for breast masses," *Acad. Radiol.* **3**(12), 1000–1006 (1996).

26. P. M. Taylor, "A review of research into the development of radiologic expertise: implications for computer-based training," *Acad. Radiol.* **14**(10), 1252–1263 (2007).
27. A. Ganesan et al., "A review of factors influencing radiologists' visual search behaviour," *J. Med. Imaging Radiat. Oncol.* **62**(6), 747–757 (2018).
28. Z. Gandomkar et al., "Recurrence quantification analysis of radiologists' scanpaths when interpreting mammograms," *Med. Phys.* **45**(7), 3052–3062 (2018).
29. C. A. Beam, E. F. Conant, and E. A. Sickles, "Factors affecting radiologist inconsistency in screening mammography," *Acad. Radiol.* **9**(5), 531–540 (2002).
30. E. U. Ekpo et al., "Mammographic breast density: comparison across women with conclusive and inconclusive mammography reports," *J. Med. Imaging Radiat. Sci.* **47**(1), 55–59 (2016).
31. T. Drew, M. L.-H. Vo, and J. M. Wolfe, "The invisible gorilla strikes again: sustained inattentional blindness in expert observers," *Psychol. Sci.* **24**(9), 1848–1853 (2013).
32. D. J. Simons and C. F. Chabris, "Gorillas in our midst: sustained inattentional blindness for dynamic events," *Perception* **28**(9), 1059–1074 (1999).
33. B. Lumbreras, L. Donat, and I. Hernández-Aguado, "Incidental findings in imaging diagnostic tests: a systematic review," *Br. J. Radiol.* **83**(988), 276–289 (2010).
34. H. L. Kundel et al., "Searching for lung nodules. A comparison of human performance with random and systematic scanning models," *Invest. Radiol.* **22**(5), 417–422 (1987).
35. D. P. Carmody, C. F. Nodine, and H. L. Kundel, "An analysis of perceptual and cognitive factors in radiographic interpretation," *Perception* **9**(3), 339–344 (1980).
36. L. Ebner et al., "Variations in the functional visual field for detection of lung nodules on chest computed tomography: impact of nodule size, distance, and local lung complexity," *Med. Phys.* **44**(7), 3483–3490 (2017).
37. M. Lago et al., "Measurement of the useful field of view for single slices of different imaging modalities and targets," *J. Med. Imaging* **7**(2), 022411 (2020).
38. D. M. Levi, "Visual crowding," *Curr. Biol.* **21**(18), R678–R679 (2011).
39. J. M. Wolfe, "Guided Search 6.0: an updated model of visual search," *Psych Bull. Rev.* 1–33 (2021).
40. 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).
41. H. Bulu, A. Alpkocak, and P. Balci, "Uncertainty modeling for ontology-based mammography annotation with intelligent BI-RADS scoring," *Comput. Biol. Med.* **43**(4), 301–311 (2013).
42. 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.: Principles Implications* **3**(1), 10 (2018).
43. J. M. Wolfe and M. J. Van Wert, "Varying target prevalence reveals two dissociable decision criteria in visual search," *Curr. Biol.* **20**(2), 121–124 (2010).
44. C.-C. Wu and J.M. Wolfe, "Useful Field of View shows why we miss the search target when we "look at" it," *J. Vision* **19**(10), 314d (2019).
45. J. Grobbel et al., "Normal values for the full visual field, corrected for age- and reaction time, using semiautomated kinetic testing on the octopus 900 perimeter," *Translational Vision Sci. Technol.* **5**(2), 5–5 (2016).
46. H. J. Godwin, E. D. Reichle, and T. Menneer, "Modeling Lag-2 revisits to understand trade-offs in mixed control of fixation termination during visual search," *Cognit. Sci.* **41**(4), 996–1019 (2017).
47. J. Wolfe et al., "UFOV Mammography Exp. 2," https://osf.io/kau3q/ (2020).
48. W. J. Tuddenham, "Visual search, image organization, and reader error in roentgen diagnosis. Studies of the psycho-physiology of roentgen image perception," *Radiology* **78**, 694–704 (1962).
49. K. S. Berbaum et al., "Satisfaction of search in diagnostic radiology," *Invest. Radiol.* **25**(2), 133–140 (1990).
50. K. S. Berbaum et al., "Satisfaction of search in traditional radiographic imaging," in *The Handbook of Medical Image Perception and Techniques*, E. A. Krupinski and E. Samei, Eds., pp. 107–138, Cambridge University Press, Cambridge, UK (2010).

51. 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).
52. D. W. Stephens and J. R. Krebs, *Foraging Theory*, Princeton University Press, Princeton, New Jersey (1986).
53. D. R. DeVries, R. A. Stein, and P. L. Chesson, "Sunfish foraging among patches: the patch departure decision," *Anim. Behav.* **37**, 455–464 (1989).
54. J. M. Wolfe, "When is it time to move to the next raspberry bush? Foraging rules in human visual search," *J. Vision* **13**(3), 10 (2013).
55. J. M. Wolfe, T. S. Horowitz, and N. M. Kenner, "Rare targets are often missed in visual search," *Nature* **435**(7041), 439–440 (2005).
56. K. K. Evans, R. L. Birdwell, and J. M. Wolfe, "If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening," *PLoS One* **8**(5), e64366 (2013).
57. J. M. Wolfe et al., "Low target prevalence is a stubborn source of errors in visual search tasks," *J. Exp. Psychol.-Gen.* **136**(4), 623–638 (2007).

Biographies of the authors are not available.