Expertise in visual search of medical and non-medical images

Ann J. Carrigan

B.Sc. (Medical Radiation Science)

Grad.Dip.App.Sc. (Medical Ultrasound)

B.Sc. Psychology (Honours)

Perception in Action Research Centre

Centre for Elite Performance, Expertise and Training

ARC Centre of Excellence in Cognition and its Disorders

Department of Cognitive Science

Faculty of Human Sciences

Macquarie University

This thesis is submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

September 2017

Table of Contents

Thesis abstract

This thesis aims to understand the crucial cognitive mechanisms that underpin visual search in medical images and the influence of expertise. Research suggests that experienced radiologists use information from an initial glance at an image to set the basis of their diagnosis. This research explores the information that can be extracted from images regarding the presence and location of a target. I use two stimulus types: real medical images on which I test both novices and expert radiologists, and natural scenes as a model for radiologist search. In Chapter 1, I present an overview of the literature relevant to these aims. Chapter 2 presents two experiments where I showed that a target could be both detected and located in a natural scene after a brief presentation (33ms), but that visual clutter interferes with performance in a predictable way. In Chapter 3, I showed radiologists were able to detect and localise an abnormality in a mammogram presented for 250ms at levels better than guessing. Crucially I demonstrated that a normal patient variant, high levels of breast density, affects performance. I conducted an in-depth analysis which emphasises the importance of considering factors such as stimulus variability, response imprecision, and participant guessing. In Chapter 4, I investigated the extent to which expertise guides attention based on prior experience with the prevalence of cancer, using a novel cueing paradigm where a chest radiograph (with or without a suspicious nodule) formed a prime. For naïve observers, an artificially boosted nodule in the prime radiograph guided attention, validating the task. Radiologists viewing true, more subtle nodules did not show the same effect, nor did they show any attentional guidance from cancer prevalence. However, more experienced radiologists seemed to be more sensitive to the subtle nodules than less experienced radiologists, suggesting that

expertise might boost nodule salience. Finally, in Chapter 5, the implications of these findings are discussed in a broader context along with suggestions for areas of future research. Overall, my research shows that there is a large amount of information available after observers first look at a scene or medical image; more than previously thought. Further, the visual complexity of the display affects performance. Together, the experiments presented in this thesis advance the scientific understanding of the type of information available in the first glance and has clear implications for radiologist teaching and clinical benefits.

Author Statement

I certify that the research presented in this thesis is my original work and that any sources of information or assistance are cited appropriately. In addition, I certify that the work in this thesis has not been previously submitted for a degree nor has it been submitted to any university or institution other than Macquarie University.

The research presented in this thesis was approved by the Macquarie University Human Research Ethics Committee (Medical Sciences) (reference number: 5201400567).

Signed,

Ann Carrigan

Acknowledgements

Often you hear that the PhD journey is an isolating one, but mine was far from it. I thrived and relished the opportunity to spend time with my peers, grow personally and academically and to absorb knowledge from other great minds both visiting academics and staff at MQ. I have loved walking through the doors each day, the collegiate atmosphere and have jumped in with both feet when opportunities were presented. I have learnt so many things including perseverance and resilience. It helps that I am incredibly passionate and motivated about my topic. Remembering this often boosted me along during the tough days.

My passion for research began sitting in a psychology undergraduate course: Cognitive Processes II, when slides came up about attention and visual search, specifically the work done by Dan Simons and Jeremy Wolfe. That was my light bulb moment. As a medical imaging professional, I realised that we still had much to learn about what goes on when we are searching complex medical images.

There is no doubt that I wouldn't be here if not for my PhD supervisors. Anina: you truly are my daily inspiration who pushes, encourages, and holds the bar extremely high. Hopefully I finish a little closer to that bar. Thank you for your mentorship, for your patience, and inspiring me to be a stronger 'woman in science'. Both working mums, I felt that you understood those days when life got in the way of my PhD. Also, that life *should* get in the way: personal well-being and family comes first. There was never a moment when you were not there for me. So, thank you. Susan: thank you for sharing this journey with me. I appreciated your wisdom and encouragement, and I always felt a sense of calm after our meetings. Thank you especially for your 'fresh eyes'.

Eugene, my psychology honours supervisor. Always encouraging and forever wise. Thank you for your mentorship, giving me the confidence to grow and take leaps off cliffs, for allowing me to tutor on statistics and also for our regular coffee catch-ups.

I consider myself very fortunate to have met, travelled to and spent time in their labs, been guided, encouraged and mentored by some of the greatest minds in the field. I strive to be a better researcher because of these people that light the path: Jeremy Wolfe, Todd Horowitz, Stephen Mitroff, Trafton Drew, Ruth Rosenholtz, Mark McEntee, Elizabeth Krupinski and Patrick Brennan.

Thank you to the radiologists who gave up their valuable time to do my studies, especially NSW BreastScreen. Also to my work colleagues at San Radiology, especially Dr Phil Scamps, Dr Mandy Ballie, Dr Phil Hung and Melinda Happer, who helped me throughout my PhD journey.

To the PARC group. Thank you for listening to my talks and providing valuable feedback when needed. Also for introducing me to the wonderful world of neuroscience. Friends both at MQ and home, thank you for providing the necessary lunch and coffee catch-ups. Cognitive science staff especially Lesley, Craig and Marcus.

My beautiful and loyal companion, Cash. Our daily walks in the fresh air allowed me the clarity to work, especially over the final months. Joy, for stepping in when I asked for help.

To my strong, loving and intelligent Mother, Elaine. As an eight-year-old, you taught me that it was possible to be a mum and continue your education. Thank you for your love, support and interest in my projects.

Emma Joy, my gorgeous daughter: you are strong, brave and an ‘old soul’ who brings me much happiness: thank you for keeping me grounded and for your spa

treatments. Rohan Gordon, my gentle, loving son: thank you for your 7-second hugs and teaching me to mountain bike. Thank you both for listening to me, for giving me the space to grow, and providing me with much needed distractions. Finally, thank you for cooking me creamy pasta dinners, sharing your chocolate (even your bilbies) and making my heart sing with pride and joy.

Matt: thank you for supporting us in every way possible while I undertook this journey, for listening to me, for your love, help, care and patience. You appreciated early on, that a learning wife, is a happy wife. Your intelligence inspires me and I love that we can now discuss topics such as data and statistics together (as well as critically evaluate the media).

Finally, for my Father, Gordon, who sadly passed half way through my PhD journey. Thank you for passing on your love of the country and ocean and providing me with many joyous memories that include body surfing at Whoopie (Woolgoolga) together. Also for instilling in me the curiosity to keep learning, your Dad jokes and for teaching me to ‘not miss the best part of the day’, to ‘not overdo it’, ‘what goes up must come down’ and ‘that’s the way the cookie crumbles’. Thank you to my team (both personal and at MQ) who pulled me through this difficult time. I feel his love gently guiding me every day.

This thesis is dedicated to him.

Author Note

This thesis has been prepared in the form of a thesis by publication; as such, there is a degree of repetition, particularly in the introduction of Chapter 4. I have tried to avoid repetition as much as possible while allowing each chapter to stand on its own. The formatting and referencing style adopted throughout this thesis reflects the requirements of the APA Publication Manual (6th edition).

Chapter 1: General Introduction

General Introduction

Radiologists are medical doctors who have undergone specific postgraduate training in performing and interpreting diagnostic imaging tests and interventional procedures. They are required to search and interpret complex medical images to reach a diagnostic decision. The result of this search forms the core component of the information for diagnostic and treatment decisions about patients. The focus of this thesis is the visual search component of medical imaging where the consequences of errors range from mild to severe. In a worst-case scenario, a cancerous mass could go undetected which could have a devastating outcome for a patient. Understanding the basic processes that underpin radiologist visual search is a crucial component in better understanding the causes and possible solutions for errors.

The tests used by radiologists involve the use of modalities which provide 2-dimensional displays (e.g., plain X-ray, mammograms) as well as dynamic (e.g., ultrasound) and 3-dimensional examinations (e.g., computer tomography and magnetic resonance imaging). According to the Royal Australian and New Zealand College of Radiologists (RANZCR: http://www.ranzcr.edu.au/radiology), to qualify as a radiologist requires (after completion of a six-year medical degree) two years' internship, then an accredited five-year training program, whilst practicing within a radiology setting. Once qualified, radiologists can choose to work in various subspecialties of radiology such as breast, interventional, musculoskeletal, cardiac, neuroradiology or paediatric imaging. Using medical imaging, radiologists form a core part of a multidisciplinary team trained to assist other doctors and specialists who treat patients by making a diagnosis and providing treatment (e.g., drug therapy). This extensive training means they have the medical knowledge and ability to understand

medical problems or symptoms by the interpretation of anatomical representations of the human body.

The task of searching medical images for abnormalities is effectively a visual search in a cluttered display. Similar to basic search paradigms, radiologists are required to visually search a cluttered medical image for a target (e.g., mass) while ignoring distractors (e.g., normal anatomical structures) and then make a diagnostic decision based on abstract anatomical features. To make a diagnosis, the image must be perceived and interpreted appropriately; these tasks are perceptually and cognitively demanding. Errors that can occur are either misses of a target that is present or false alarms on target absent displays. Coupled with the increasing amount of computer interaction, 3D imaging modalities and a higher caseload, the potential for error is high (Samei & Krupinski, 2010). All tasks that involve visual search are prone to error and, similar to other real world visual search tasks, radiologist errors may result in serious consequences for public safety.

The visual search errors we know humans are prone to make become particularly problematic in such high consequence environments as radiology. The consequences for making an error can be extremely serious: false negatives result in missed abnormalities, which can be catastrophic. For instance, a missed breast carcinoma on a mammogram means delayed treatment and thus reduced survival rate. It has been reported that for radiologists there may be up to a 30% miss error rate and an equally high false alarm rate (Berlin, 2005). Using eye tracking methods, Kundel, Nodine and Carmody (1978) categorised radiologist errors when reading chest radiographs into the following: visual search errors, where they never fixate the abnormality (30%); recognition errors, where the abnormality is fixated but only briefly (25%); and decision errors, where the abnormality is fixated but actively dismissed as

an abnormality (45%). Other error types include satisfaction of search (subsequent search misses), where detection of the first abnormality interferes with detection of others (Adamo, Cain & Mitroff, 2013; Berbaum, et al., 1990; Tuddenham, 1962) and inattentional blindness, where sustaining attention on a specific target can cause an obvious, unexpected target to be missed (Drew, Võ, & Wolfe, 2013; Mack & Rock, 1998; Simons & Chabris, 1999). Furthermore, a robust phenomenon reported in both the vision science and medical literature is the effect of prevalence, where low target prevalence (i.e. rare targets, few cases are truly abnormal) results in elevated miss rates (Evans, Birdwell & Wolfe, 2013; Mitroff & Biggs, 2014; Wolfe, Horowitz & Kenner, 2005). Given much radiology search is within a screening environment, where the prevalence of an abnormality is very low, the effect of rare prevalence is likely to contribute to high rates of miss errors. The evidence that errors clearly occur, and the cost of such errors in both financial and social terms, makes it crucial to understand the processes underpinning visual search in medical images.

Anatomically speaking, the human body is complex and variable: both externally (e.g., height) and internally (e.g., breast tissue), there is considerable variation between individuals. In mammograms, the breast parenchyma is highly variable with regards to level of breast density (mammographic breast density: MBD). Dense tissue is comprised of normal fibroglandular tissue and appears radio-opaque on a mammogram, whereas tissue that is comprised of fat appears radio-lucent. In practice, radiologists use density to characterise the complexity of the breast parenchyma (Ray, et al., 2016). It has been shown that as MBD increases there is an increased risk (4-6 fold) of breast cancer (Boyd, et al., 2010). This is most likely due to the combined effects of lower radiologists' sensitivity when the mammogram is complex; this could mask and/or distract from existing pathology, thus reducing accuracy; and with the

increased amount of breast tissue present in a dense breast in which a cancer could occur. These combined effects negatively impact a patient as they can limit early detection of breast cancer and subsequent treatment (Al-Mousa, Ryan, Mello-Thoms & Brennan, 2014).

The focus of this thesis is on the cognitive mechanisms that underpin visual search in medical images by radiologists. Such searches, however, share key characteristics with laboratory visual search tasks, which have been well studied for decades. I therefore start my literature review by outlining the factors that are known to strongly impact on visual search. I then focus in on previous medical imaging literature before outlining my major research questions and the structure of the thesis.

1.1 Overview

The moment we wake up each day, we are required to process a large amount of sensory information in order to perform tasks and attain our goals. Looking for a target amongst distractors is an essential and often challenging task that we routinely conduct throughout our daily lives. In order to find a coffee cup or our car keys, we must engage our attention, actively scan our environment, whilst ignoring irrelevant items in our surrounds. Object recognition is capacity limited, so we need to be able to search and direct our attention in order to function (Wolfe, Võ, Evans, & Greene, 2011). Selective attention is one way we can filter the incoming information, and is actively engaged when conducting visual search tasks (Lamy, Zivony, & Yashar, 2011).

Recent work in visual search has linked decades of laboratory studies using perceptually simple displays (e.g., Posner, 1980; Treisman & Gelade, 1980; Wolfe, 1994) with real-world search tasks across a wide range of domains. There are many

professionals who rely upon visual search to effectively carry out their routine work. These include airport luggage screening, where images of luggage are searched for weaponry and prohibited items (Wolfe, et al., 2005), air traffic surveillance and aeroplane piloting, where ever-updating and moving displays are monitored (Lopez, Previc, Fischer, Heitz and Engle, 2012), and central to the theme of this thesis, medical imaging, where a radiologist searches an image or scan for abnormalities (Drew, et al., 2013; Evans, Georgian-Smith, Tambouret, Birdwell & Wolfe, 2013; Evans, Haygood, Cooper, Culpan, & Wolfe, 2016; Kundel & Nodine, 1975; Kundel, Nodine, Krupinski & Mello-Thoms, 2008). Studying real-world visual search is crucial as even the experienced observer has perceptual limitations and a better understanding of these limitations will help improve training and practice of these tasks.

Evidence from both the natural scene and medical imaging literature has demonstrated that a large amount of information is processed in the first glance at a scene (e.g., Kundel & Nodine, 1975; Potter, 1976, Thorpe, Fize & Marlot, 1996). Understanding this ability is critical as important decisions (e.g., medical diagnosis) often depend on this early processing. An established theory in the medical perception literature proposes that the information extracted from the initial signal, that indicates an abnormality, guides subsequent search to the location (Kundel & Nodine, 1975). However, recent studies report that this signal does not guide the observer to where an abnormality is located, but rather changes the observer's search strategy (Evans, et al., 2013; Evans, et al., 2016). This recent challenge to the classic model of radiologist search deserves full consideration, as it has important theoretical implications.

The primary aims of this research are to investigate and understand the processes that underpin important diagnostic tasks related to medical images and to explore visual search expertise in the initial stages of perception. To this end, my

research comprises three series of experiments exploring the information that can be extracted from images regarding the presence and location of a key target (e.g., a mass in a mammogram). Specifically, I explore the time course of target detection and localisation (in both natural scenes and medical images), the effect of the visual complexity of a display, and the mechanisms that drive attention. In this introductory chapter I start by taking a step back by discussing visual attention and search and look at some of the factors that influence the allocation of visual attention. Then, I introduce some of the models based on other stimulus sets. I review studies exploring fast visual processing and expertise where I specifically focus on object detection and localisation (in scenes) and abnormality detection (in medical images). Next, I introduce the dominant models of radiologist visual search. I then return to some of the specific issues in radiology that I will be addressing in this thesis. I conclude this chapter by examining recent findings that raise some challenges to how a radiologist searches medial images that are driving some of the major research questions in this thesis and outline how I have approached these questions in this thesis.

1.2 Visual attention and search

Attention is capacity limited: we simply cannot attend to and process everything that arrives at our visual system at any one time. Many theorists have described a 'bottleneck', where information arrives in parallel but only the relevant information is filtered through for processing (although there is debate about the stage at which this filtering occurs; Broadbent, 1958; Treisman, 1964). The factors which drive our attention act to select and prioritise this information for processing. There is a dynamic interplay between what is happening in the environment and our internal state: attention

can be guided by bottom-up (exogenous) information from the stimulus as well as by top-down (endogenous) information about one's goals (Jonides, 1981; Posner, 1980). The interaction between exogenous and endogenous influences on attention determines what information passes through into the limited-capacity system. The research in this thesis relates to spatial attention, the directing of attention to a location in space to select and prioritise relevant information.

Visual search is an excellent paradigm for measuring the effect of different factors on the guidance of spatial attention. A classic example that illustrates many of these factors is the well-known puzzle, *Where's Waldo?* (Hanford, 1987; Figure 1). The target of the search is a character that is comprised of the unique combination of certain features such colour and shape; Waldo wears a red striped shirt and blue jeans. Many other items in the display share these features, which makes the task difficult. This illustrates target-distractor similarity, a factor that effects the degree to which target features can be used to guide attention efficiently. Prior research has shown that target salience is higher and search efficiency increases when targets are dissimilar from the distractors in a display (Duncan & Humphries, 1989; Koch & Ullman, 1985; Koehler, Guo, Zhang, & Eckstein, 2014). Other factors that increase the difficulty of finding Waldo are the complex scenes with many heterogeneous objects. Research with perceptually simple displays has demonstrated the strong effect that increasing the number of distractors can have on searches for targets that are not defined by a unique feature (setsize: Treisman & Gelade, 1980; Wolfe, 1994), distractor heterogeneity (Duncan & Humphries, 1989), and additional visual information (clutter: Rosenholtz, Li, Mansfield & Jin, 2005). These factors have been studied more systematically in classic laboratory visual search tasks which I will now review.

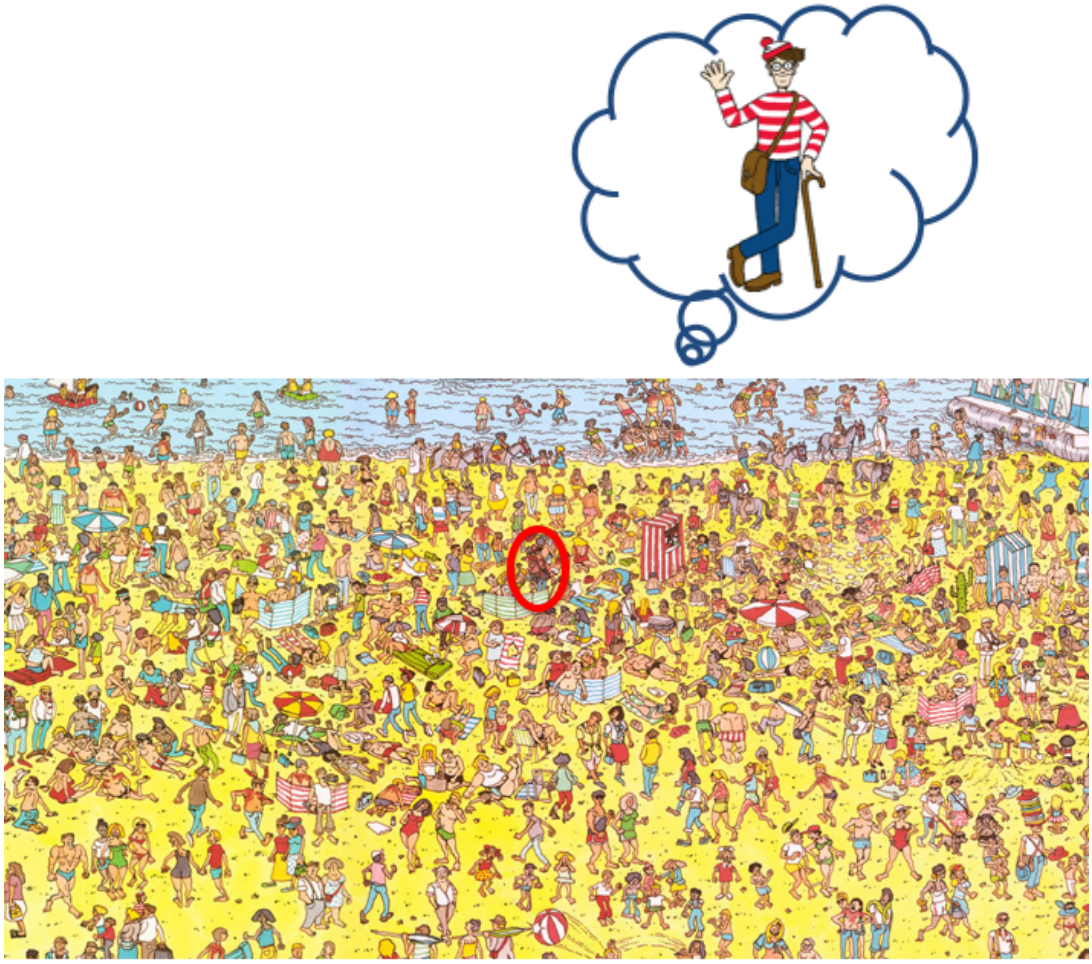


Figure 1. Example of the pictorial visual search puzzle Where's Waldo (Downloaded from Google images). Note: the red circle surrounding Waldo is for illustration purposes only.

In vision science, visual search is studied experimentally in a laboratory setting using well-controlled displays and computer-based tasks. In a typical visual search paradigm, observers are asked to find a target among distractors in a visual display that stays on until response (free viewing); usually 50% of displays contain a target. Measures include accuracy and reaction time to target detection report. Figure 2 shows

a basic visual search array (for target-present displays) and illustrates key factors that make visual search difficult.

Based on decades of visual search data, we know that some features result in 'pop-out' search. Stimulus features, such as unique colour and orientation, can guide attention in a bottom-up manner (Posner, 1980). Figure 2a demonstrates an array where the target (red letter L) differs from the distractors (black Ls) by a single feature. 'Pop-out' or highly efficient search is not affected by the number of distractors in the array. Stimulus driven or bottom-up guidance will attract visual attention to certain aspects of a display or scene, resulting in a highly efficient search.

Attention and search can also be guided by top-down factors, which allows search in situations where the target does not 'pop-out'. Figure 2b shows an example of a type of serial search where attention has to be allocated item-by-item to find the target (black T) among the distractors (black Ls) (Triesman & Gelade, 1980; Wolfe, 1994). Search for a target defined by a conjunction of features tends to be slower than feature search and is affected by the number of distractors (Treisman & Gelade, 1980; Wolfe, 1994: Wolfe & Horowitz, 2017). There can, however, be guidance to some features, which increases the efficiency of search over a truly serial search (e.g. feature-based guidance to red, Figure 2c. For target absent displays, conjunction search is slower than target present displays, as it takes more time to actively search all the items instead of stopping once the target is found, which takes, on average, a search of half the items in the display (Wolfe, 2012). There have been many hundreds of studies using this basic paradigm to explore the way different factors influence visual search efficiency.

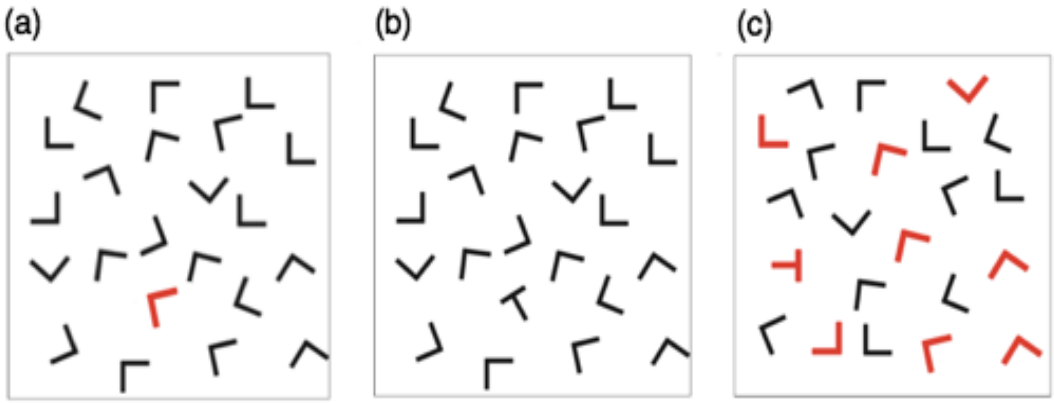


Figure 2. Basic visual search paradigm illustrating target present displays. (a) Illustrates a 'pop-out' search: the target (red L) varies on a unique feature.; (b) Illustrates a type of serial search: the target (the letter T) is presented among many distractors (Ls); (c) Illustrates guided search using feature based attention: the target is a red T (adapted from Wolfe & Horowitz, 2017).

There are a number of key factors that affect the speed with which a target can be located. Top-down attentional control settings can be quite complex and include feature guidance, scene guidance, prior knowledge and item value (Wolfe & Horowitz, 2017). In feature guidance, when you are looking for a particular target, features of that target are boosted in a top-down manner that can guide attention (Jonides, 1981). For example, searching for a target you know is red results in feature-based attentional enhancement of neural responses to items throughout the visual field that share the target colour, providing a mechanism for guided search (Wolfe, 2012). Actually, even search for a target defined by a unique feature includes a top-down component: the task is to detect that salient target, making it task-relevant and therefore presumably

facilitated by both bottom-up and top-down guidance. Figure 3 provides an example of targets that evoke top-down feature and bottom-up guidance in a visual search.

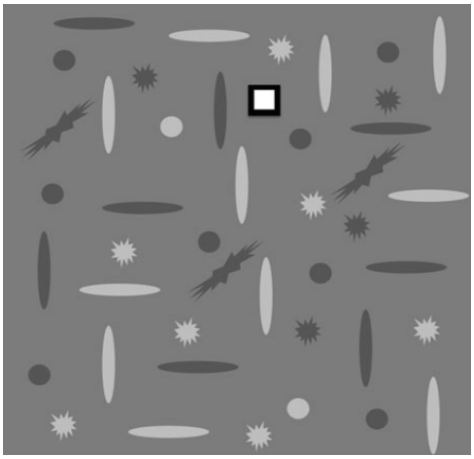


Figure 3: Illustration demonstrating bottom-up and top-down guidance. If the target is the white square with a black border set among grey circles and shapes, it can be found using bottom-up guidance as it is highly dissimilar to the distractors. If one is searching instead for the light grey circle, top-down guidance can be used to restrict search to the sub-set of light grey round-ish objects (from Wolfe, Evans, Drew, Aizenman & Josephs, 2016).

Folk, Remington & Wright (1994) describe situations where attentional control settings can be set or 'tuned' to specific features such as colour, motion or onset. They based this on results of experiments in which observers were instructed to attend to a critical feature of the target (such as colour). They reported that under these circumstances other features (such as motion or onset) failed to capture attention. This 'colour set' can also be set to within-colour dimensions, meaning that if red is the critical feature then this will capture attention and other colours can be filtered out (Folk & Remington, 1998). These results could also be considered within a general

similarity account: distractor items that are more similar to the target capture attention more than items less like the target (Duncan & Humphries, 1989). For example, search for an upright L among 180° and 270° rotated Ts is slower compared with 0° and 90° rotated Ts, due to the similarity in the length of the letter strokes in the 180°/270° condition (Duncan & Humphries, 1989). More recently, Becker (2008) proposed that a relational set or contextual properties of the target act to guide visual attention in a top-down manner. In a series of attention cueing experiments, attention was shown to be tuned to relational properties of the target (e.g., smaller) when compared with the distractors. Furthermore, that knowledge about these relations develop with experience (Becker, Folk & Remington, 2010). These findings highlight how fine-tuning of attention to features to match task requirements, and previous experience with a task, can both influence the allocation of attention.

Another important factor that has been added to attentional models of visual search is scene-based guidance. This is based on the idea that attention allocation is influenced by our prior knowledge about the spatial relations of objects in scenes. Natural scenes are complex but the items within them are not randomly located. We have a rich and extensive experience with scenes; for example, rules, knowledge and experience has taught us that a toaster is usually located on the kitchen bench (Chun & Jiang, 1998; Wolfe, et al., 2011). Evidence of top-down, scene structure guidance can be found in the scene congruency literature which has shown that observers are faster at scene categorisation when the objects contained within the scene are consistent with the surrounds (Davenport & Potter, 2004; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007). For instance, a car on a highway is consistent with our expectations, whereas a car in the sky is not. In the example of natural scenes, visual search is driven by searching for

the most distinctive features of a target in the locations of scenes most likely to contain the target, based on prior experience of similar scenes (Wolfe et al., 2011). Thus, when we move beyond simplistic displays, additional sources of guidance come into play.

Prior experience with specific search displays can also influence how attention is guided (Wolfe & Horowitz, 2017). A number of studies describe search facilitation by repetition, such as implicit priming for previously seen stimuli (Becker et al., 2010; Kristjánsson, 2006; Lamy, et al., 2011). Priming occurs when search improves (faster responses) on a display when the location of distractors and the target is repeated over trials, despite the participants being unaware of this. Chun and Jiang (1998) describe this search facilitation as contextual cueing. In a series of well-designed studies with tight experimental control, observers were presented a search display with varying spatial layouts of targets (Ts) among distractors (Ls). Half of the layouts were repeated across the experiment with the spatial location of the target constant, but observers were unaware of this repetition. The findings showed that on the repeated trials the target was localised and discriminated faster suggesting that target context (spatial relations) was implicitly learned over the course of the experiment. These experiments show that attention can be guided and responses biased by top-down factors based on previous exposure.

Further evidence for how priming influences search and attention also comes from studies using attentional cueing paradigms. Here a lateralised cue is specified in a spatial location preceding the subsequent display that contains a search stimulus. The previously seen prime acts to shift the observer's attention to the location of the target. Attention can be shifted by exogenous cues (e.g., salient prime) and endogenous cues (e.g., informative prime). When an observer is primed or cued by stimuli, subsequent reaction times (RTs) to a target in that same location (valid trials) are shorter than

targets in different locations (invalid trials). Performance is also facilitated for validly-cued compared to uncued or neutrally-cued trials (Becker et al., 2010; Posner, 1980). Priming could potentially occur in number of real world situations that could affect the performance on important tasks such as medical screening. In Chapter 4, I have investigated priming on both naïve and expert observers (radiologists) using medical images as it is important that we understand this aspect of attention guidance in applied contexts.

There have been several studies that show that voluntary attention and search can be influenced by reward and item value (e.g., Della Libera, & Chelazzi, 2006). For example, in Anderson, Laurent & Yantis (2011), observers searched for red and green targets interspersed between non-coloured targets. At the end of each trial, feedback was provided along with the accumulation of a monetary reward for correct responses. If the observers were rewarded more for red items, then attention was guided to red, even when the task was to find a different target. In addition, it has been shown that this type of learning persists over time (Anderson & Yantis, 2013). Perhaps in an applied setting, such a medical imaging, finding a cancer and thus preventing disease progression is reward alone. However, this intriguing aspect of how attention can be guided is beyond the scope of this thesis.

In summary, there are a number of factors that drive our selective attention from both the environment and our internal state. These include bottom-up and top-down factors which interact and lead us to attain our goals. These have been studied extensively in laboratory settings using typical visual search displays (e.g., Chun & Jiang, 1998; Duncan & Humphries, 1989; Folk et al., 1994; Koch & Ullman, 1985) and natural scenes (e.g., Wolfe, et al., 2011) but less so in medical imaging. It seems reasonable that the basic cognitive underpinnings of attention operate in the same way

regardless of whether you are viewing a simple display, a natural scene or a medical image, but there are certain factors that may be more or less influential in different display types, and these could be domain specific. Thus, it is important to test how the knowledge gained in highly controlled laboratory stimuli generalise to more complex medical images.

1.3 Models of visual search

Early models of attention proposed two types of visual processing that occur in parallel (Neisser, 1967). Treisman and Gelade's (1980) influential Feature Integration Theory (FIT) proposes that a first preattentive stage is automatic, unconscious and effortless, where features available in parallel across the display are 'mapped'. As reviewed in section 1.2, basic features (e.g., colour or orientation) seem to be able to be detected without attention and will 'pop out' because they differ in a basic feature from their surroundings (e.g., red circle among blue circles). A second stage applies focused attention to detect a combination of features across feature maps ('conjunctions'; e.g., the conjunction of colour and shape required to detect a red circle among blue circles and red squares). This type of search occurs serially (item-by-item) and targets take longer to be detected when set size increases. FIT (Treisman & Gelade, 1980) was fundamental in providing an understanding of how attention is allocated. These early influential models of attention provided a structure to understand how we attend to relevant items in our environment and laid the foundation for the development of further models.

All search tasks require attention to identify a target, and it has been proposed that the information extracted in the first stage can be used to 'guide' search leading to

identification (Wolfe, 1994). Originally developed over twenty years ago, Wolfe, Cave and Franzel (1989) proposed an alternative model to FIT (Treisman and Gelade, 1980), leading to Wolfe's (1994) influential model of Guided Search. This model extended the two-stage architecture of FIT (Treisman & Gelade, 1980) by adding the core principle that information from the first stage could guide attentional deployments in the second stage. Basic visual attributes, such as target size and orientation, could be used by other processes to constrain attentive search to likely target locations. This model was subsequently built upon in Guided Search 4.0 (GS4: Wolfe, 2001). The architecture of GS4 describes a non-selective pathway which performs a limited analysis of the entire scene in parallel, the contents of which subsequently guides attention. The type of guidance the meaning of a scene provides is 'scene-based' guidance. Occurring early in the visual system, this simplistic pathway allows for the rapid global assessment of an entire image and the processing of scene gist. Basic features such as orientation or the distribution of spatial frequencies can be processed, but it lacks the precision required for object recognition. A selective pathway recognises objects based on a limited subset of the stimuli. GS4 (Wolfe, 2001) has been applied to natural scenes into the 'Two-pathway architecture of visual processing' (Wolfe, et al., 2011). Figure 4 (from Wolfe et al., 2011) illustrates this model. The non-selective pathway is similar to preattentive processing providing a route to semantic scene information, but instead it occurs in parallel with the selective pathway and together they result in a successful search.

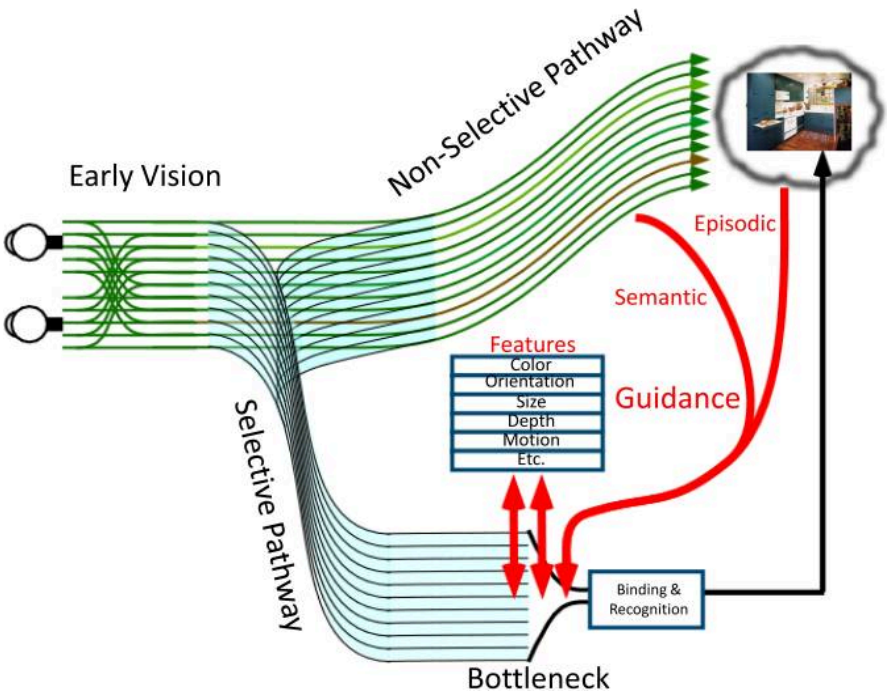


Figure 4: A two-pathway architecture for visual processing. The non-selective pathway can extract scene statistics but lacks the precision for object recognition. The selective pathway can bind features and recognise objects. Scene guidance is also incorporated into the model (from Wolfe, et al., 2011).

Although the non-selective pathway does not support object recognition, it may provide estimates such as spatial layout of the scene and how the items contained within it are distributed. This may provide a possible explanation for how the visual system is capable of tasks such as rapid detection of masses within medical images by experts (e.g., Evans et al., 2013; Evans et al., 2016) and pathology slides (Houghton, Smoller, Leonard, Stevenson & Dornan, 2015). These estimates guide the resources of the selective pathway allowing the identification of relevant features which are bound and recognised leading to precise object recognition (Wolfe, et al., 2011). Overall, these models describe the framework for the mechanisms which guide the deployment of

attention and might be useful for understanding applied visual search particularly for those whose task is to interpret medical images.

1.4 Fast visual processing and expertise

From the models of search reviewed above, it is clear that in the early stages of vision a large amount of information is processed rapidly. An exposure duration of 100ms is sufficient for observers to extract the basic meaning of natural scenes (e.g., indoor versus outdoor; Potter, 1976). When stimuli are backward masked, a duration of 20ms is enough to distinguish between scene categories at the superordinate level (e.g., man-made versus natural) and basic level (e.g., coast versus city) (Greene & Oliva, 2009; Joubert, et al., 2007). It is widely accepted that rapid scene categorisation is based on a global summary or 'gist' (Oliva, 2005). Described as the earliest meaningful stage of scene perception, after or during a glance, gist captures the global properties and overall spatial layout of a scene (Oliva, 2005; Torralba, Murphy & Freeman, 2010). These properties are based on statistical and structural cues in the scenes and stimulus based information such as the low-level features within the scene (e.g., orientation and size). Objects can also be detected at brief durations: when primed with a predefined target category (e.g., animal or truck), detection has been reported at exposures of 20-25ms, albeit unmasked (Thorpe et al., 1996; VanRullen & Thorpe, 2001) and observers tend to extract low-level visual information such as size, motion and orientation rapidly (Hidalgo-Sotelo, Oliva & Torralba, 2005; Greene & Oliva, 2009; Wolfe, et al., 2011). There is also evidence that objects and their surrounds are processed simultaneously (Davenport & Potter, 2004). After 80ms exposure, detection is facilitated when the object is semantically related to or consistent with the scene background (Davenport &

Potter, 2004). The opposite is also true: we are faster to categorise a scene if the objects within are consistent (Davenport & Potter, 2004; Joubert et al., 2007). These studies provide support for the notion that observers are processing the scene globally and are able to use their contextual knowledge with very brief presentations when making rapid decisions.

It has been proposed that the ability to extract information rapidly from a scene is the result of our experience with our environment (Drew, Evans, Võ, Jacobson & Wolfe, 2013; Wolfe, et al., 2011). Simply due to our day-to-day interaction with our surrounds, natural scenes are stimuli in which we are all expert. This 'expertise' means we may have fined tuned our visual perception skills after years of interacting with our surrounds, which supports the rapid processing of scenes.

In a number of domains, it has been suggested that experts develop perceptual and cognitive skills that are tuned to their task (Nodine & Krupinski, 1998), and where a superior ability to encode large scale visual patterns develops (Drew, et al., 2013). Studies in the chess playing literature provide an excellent example of this type of expertise: experienced players (but not novices) are able to recognise and recall chess positions in a glance (Charness, 2014). After seeing a 5 second layout of pieces on a board, master chess players can accurately recall the test-game positions far better than less skilled players. Interestingly, if the chess pieces are randomly placed around the board, analogous to a visual search display, this performance is reduced. It has been proposed that experts can extract information by grouping the features of the layout rather than at the individual level (Chase & Simon, 1973). This suggests that experts use their fine-tuned perceptual and cognitive skills to process the layout/scene globally to reach a decision. These studies show that observers who have expertise in a

particular domain can extract a large amount of relevant information from a display in a short period of time.

I now move to review the body of research that suggests an experienced radiologist can detect an abnormality after viewing images for less than one second (Kundel et al., 2008; Kundel & Nodine, 1975; Evans et al., 2013; Evans et al., 2016). One must note here that no one would expect a radiologist to base their diagnosis on this brief glance and this is not the typical way radiologists read images in clinical practice. There are other image projections, previous imaging and clinical history available to a reporting radiologist who would conduct a review under free-viewing conditions. However, these studies can provide valuable insights into the type of information available in this initial 'global' signal, and thus early visual processing.

As reviewed above, experts are able to rapidly extract meaning from a display by processing what they see globally (e.g. Charness, 2014). It is plausible that a radiologist who has experience in viewing medical images is using holistic or global processing to make diagnostic decisions after seeing an image briefly. Kundel and Nodine (1975) showed that experienced radiologists were 70% accurate in interpreting chest radiographs after seeing the image for only 200ms. Studies recording eye movements have shown that expert radiologists fixate faster and more accurately on an abnormality in mammographic images than less experienced observers, and use fewer eye movements to do so (Kundel & La Follette, 1972; Kundel & Nodine, 1975). Kundel et al. (2008) showed that within 300ms on average, mammographic readers fixate upon 67% of breast cancers. Others have shown that in around 1-2 seconds of image presentation experts fixate on true abnormalities and the majority of their subsequent scanning is to confirm that there are no other lesions. This follow-up takes about 5-10 seconds after initial fixation, after which a diagnostic decision is reached (Mello-

Thoms, et al., 2005). These studies provide evidence that an enormous amount of detail can be extracted in the first glance, especially for those with domain specific expertise such as a radiologist.

Like many other visual disciplines, medical image interpretation not only requires years of training, viewing hundreds of images, but also ongoing practice. For example, in clinical practice, breast radiologists who report mammograms in screening programs interpret 2500-4000 cases/year (Rawashdeh, et al., 2013). As expertise evolves, Kundel and Nodine (1975) propose that a radiologist shifts the perceptual mechanisms used in image interpretation from recognition-by-parts (feature-based) to the holistic recognition-by-whole. Over time, after viewing hundreds of images, it has been hypothesised that a type of perceptual fine-tuning for relevant features occurs (Nodine & Krupinski, 1998), which is similar to what has been reported in other domains. In the case of an abnormal finding, it is thought that radiologists recognise larger perceptual units as chunks and their initial response is based on a global pattern. The 'patterns' extracted in the first glance contain relevant information that is then compared with a normal anatomical template, stored in memory (Nodine & Mello-Thoms, 2010). Perceptual expertise is a learned skill that radiologists develop over time and with experience. Understanding the processes involved in reaching 'expertise' could help inform trainee radiologists.

There is evidence that the information obtained in a single glance varies according to experience (Nodine, et al., 1999). Recently, others have also shown that experienced radiologists outperform novices on abnormality detection in mammograms (Evans et al., 2013) providing further support for domain specific expertise in radiology. Figure 5 shows the typical scanning pattern of both an experienced breast reader (a) and a trainee (b) (from Kundel & La Follete, 1972). This shows the efficient

search pattern of an expert who adopts a ‘look – detect – scan’ of the mammogram. The abnormality is fixated upon quickly, with higher accuracy and with less fixations. It may be that the expert is able to use the initial or holistic representation to allow for this accuracy and efficiency. This could occur as the result of years of training and perceptual learning about abnormal features (the target) and the normal background in the surrounding breast tissue. The less efficient search pattern for a trainee shows they adopt a ‘scan – look – detect’ strategy, making more fixations and spending more time searching the entire image (Kundel et al., 2008). These studies show that experience facilitates search efficiency.

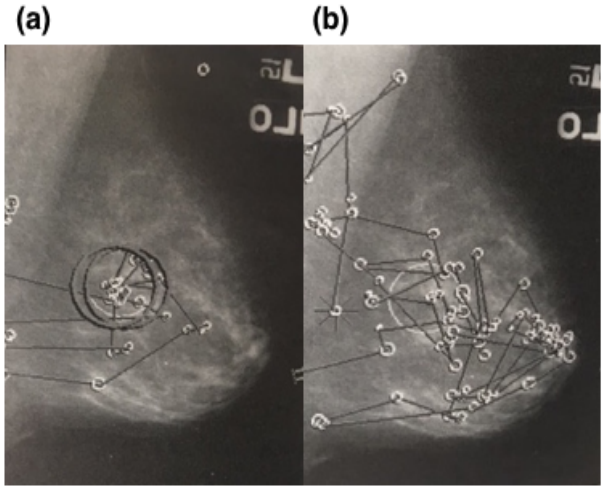


Figure 5: Mammograms showing the typical search patterns of radiologists. (a) An experienced radiologist who adopts an efficient search strategy; (b) A trainee radiologist who adopts a less efficient search strategy. Each small circle shows a fixation and the large circles represent the region of abnormality (from Kundel & La Follete, 1972).

As reviewed above, radiologists extract a large amount of information in the early stages of visual processing. For an experienced radiologist, there is some evidence that what is processed in the first second influences the overall diagnosis. In a study that investigated expert radiologists, Mello-Thoms (2009) presented two view digital mammograms (cranio-caudal and medio-lateral) and asked the radiologist to make a detection response (is there a mass or not?) followed by a localisation response (click on the mass) in both views under free viewing conditions. The accuracy of the initial localisation decision affected their subsequent decision. For instance, if the first decision was a true positive (clicking on a mass), the probability they would make a further correct decision on the second view was 94%. If the first decision was incorrect (e.g., false positive), the probability that the radiologists would make a further correct decision was only 6%. Mello-Thoms (2009) suggests that the observer developed a 'blindness' to the features of a true lesion and suggest that the first response reflects holistic/global processing without focused search. However, it is important not to place too much emphasis on this interpretation as the viewing time was unconstrained and there was no specific manipulation ensuring that this was not indeed an attentive search. Moreover, this study used a small sample size ($n = 4$) and only one classification of breast pathology. This study proposed the first decision made affects the subsequent decision, albeit at longer viewing durations. The claim that radiologists extract information in the first second that influences diagnosis, together with the evidence from the basic vision literature about the enormous amount of visual processing that occurs in the first glance, highlights the importance of understanding these initial stages of visual search.

1.5 Models of the stages of radiologist visual search

Over 40 years ago, Kundel and Nodine (1975) proposed a model of the stages of radiologist visual search that has become a well-accepted model in the medical imaging field. This model describes two distinct stages leading to a diagnostic decision. During the first, a global impression or 'gestalt' of the image is compared with 'normal' information, or a schema, stored in long term memory. This schematic representation is thought to have developed as a result of viewing large numbers of negative cases relative to abnormal cases (Kundel & Nodine, 1975). If a perturbation is noticed, this information constrains and guides subsequent search in the second stage to the region of an abnormality. In this model, location information must be present in the initial view in order to guide search (Kundel & Nodine, 1975; Nodine & Mello-Thoms, 2010). Figure 6a shows the Kundel and Nodine (1975) model of visual search, illustrating that the initial percept establishes content which guides subsequent search.

Recently the Kundel and Nodine (1975) model has been challenged by Evans and colleagues (2013; 2016). These authors argue that the initial signal may alert the radiologist with an overall 'hint' or 'hunch' that there is an abnormality. This alert, rather than guiding search to its location, changes the search strategy to a more thorough search. This signal could be supported by the rapid extraction of basic information such as the summary statistics of an image, within the non-selective pathway in the 'two-pathway architecture for visual processing' (Wolfe et al., 2011) (as reviewed in section 1.3). Evans and colleagues (2013; 2016) propose that this pathway could signal an abnormality, but fine detail, such as its location, becomes available along the selective pathway at a later period.

Figure 6b shows the proposed Evans et al. (2013) model illustrating this change in search strategy. In the next section, I review the studies that form the basis for this

model in detail, as they are the starting point of my own research. To summarise, however, they conducted a series of experiments where radiologists viewed rapidly presented mammograms (250ms – 2000ms) and were asked to make a detection and localisation decision. They found an apparent dissociation between the ability to detect a target and the ability to locate it. Although the authors acknowledge that in some cases the initial signal will guide the deployment of attention, they argue that a dissociation between abnormality detection and localisation suggests that detection of an abnormality could be supported by a global/holistic signal. As I discuss in the next section, the key issue is whether or not Evans et al. (2013) and Evans et al. (2016) have provided good evidence of such a dissociation. The results presented in this thesis inform this debate and I return to these models in the general discussion (Chapter 5).

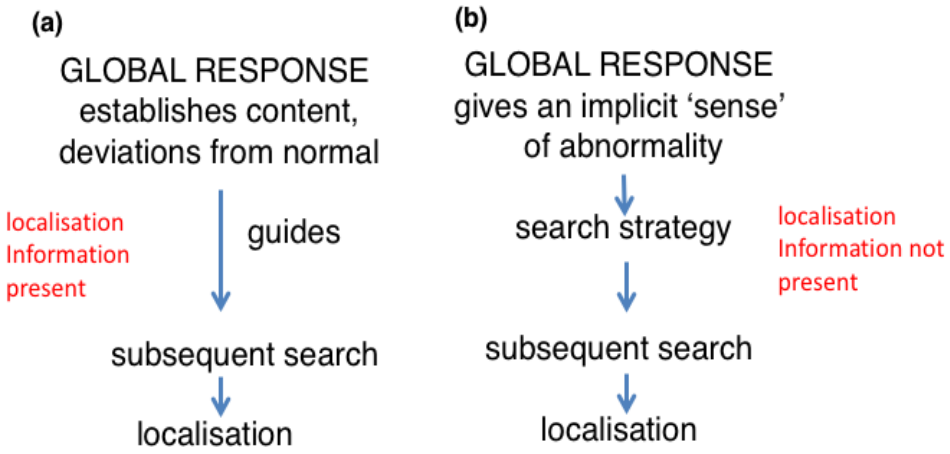


Figure 6. Basic representations of models of the stages of visual search leading to diagnosis in radiology. (a) Kundel and Nodine (1975); (b) Evans and colleagues (2013). The critical difference is whether the global response/signal contains localisation information (Kundel & Nodine, 1975) or not (Evans et al., 2013).

1.6 Rapid target detection and localisation

When we glance at a scene, we are not just interested in *what* is there, but also *where* each object is so that the appropriate action can be made. There is a large body of literature on target/object detection after fast presentations, some of which I have reviewed above, but less is known about the degree to which information about location is available in brief displays. Although it seems intuitive that when we see something we can also say where it is, there are suggestions in the change detection and medical imaging literature that detection and localisation do not necessarily go together (e.g., Evans et al., 2013; Howe & Webb, 2014). It is important to find out whether localisation and detection can be dissociated as this has theoretical implications and might inform visual search models both in general, and the models of diagnostic process from visual search of medical images in particular.

Compared with target detection, localisation has scarcely been investigated in rapid presentations, so we know much less about this. However, within the change detection and medical imaging literature there are suggestions that target detection and localisation are dissociable (Evans, et al., 2013; Evans, et al., 2016; Howe & Webb, 2014; Rensink, 2004). In a change detection paradigm, Howe and Webb (2014) showed observers a photograph of a face for 1.5 seconds, followed by a 1 second blank, and then another version of the same photograph with a single changed feature (e.g., removal of glasses). The task was to indicate if a change had occurred and if so, to select the change from a list of nine possible options. The results showed that a change could sometimes be detected in the absence of identification. Howe and Webb (2014) then considered whether these trials could be simply the result of correct guesses. They tested the influence of possible observer response bias by deriving an equation to explore observer guessing. This calculates the degree to which such trials (detection in

the absence of identification) could occur by a 'hypothetical observer' with no ability to detect changes without also being able to identify them. This model of observer behaviour did not provide an adequate account for the proportion of trials in which detection occurred in the absence of identification. The authors therefore concluded that the results cannot be explained by guessing suggesting there is apparent lack of information about the identity of the changes. There are two alternatives here. First, there truly was *no* identification information in the presence of true detection (a dissociation between detection and localisation). Alternatively, the task *failed to measure localisation* with sufficient sensitivity. As the authors acknowledge, low precision of the localisation response could be a possible explanation for the apparent 'lack of identification'.

There are a number of studies in the change detection literature which have explored the relationship among detection, localisation and identification. Fernandez-Duque & Thornton (2000) describe implicit localisation where localisation is above than what is expected by chance, even though the observers were unaware of the change. Others argue that that implicit change detection can guide attention to the change (Smilek, Eastwood, & Merikle, 2000). Following a series of experiments, Mitroff, Simons and Franconeri (2002) argue that successful change detection requires that the observer's attention be allocated to its location both before and after the change. In a typical flicker task, observers shift their attention from item to item searching for the change. Here, they argue that attention is not guided by the actual change *per se* rather the properties of the presented scene (e.g., salient regions, Rensink, O'Regan & Clark, 1997). These studies challenge those which have argued that detection and localisation are dissociable (e.g., Evans et al, 2013). Determining whether detection is accompanied by localisation is a key focus of this thesis.

The other key set of studies suggesting a dissociation of detection from localisation was conducted with medical images. Evans et al. (2013) and Evans et al. (2016) showed that when observers were shown mammograms for 500ms, radiologists (but not novices) could discriminate normal from abnormal images but localisation performance was consistently poor. Analogous to Howe and Webb (2014) the authors interpret their findings as implying that target detection is possible without localisation under some circumstances. In Evans et al. (2016) radiologists viewed target-present and target-absent mammograms for 500ms where the target-present images contained subtle masses and architectural distortions. The results showed that mean *d′* for detection was 1.16, significantly above chance (0), whereas localisation accuracy was not significantly greater than that expected by chance (6%). In subsequent experiments, Evans et al. (2016) suggested that diagnostic judgements could be made based on a whole and cut out patches from thus far normal mammograms where the contralateral side contained an abnormality. From this, they suggested that a global signal spread across the entire breast could support detection without carrying the information needed for localisation.

There are several issues with the interpretation of the Evans et al. (2013) and Evans et al. (2016) results. First, they used frequentist analyses and their main result rests on a null effect. This can only show whether performance was significantly above chance, not whether there is a dissociation between detection and localisation. Second, their conclusions were based on summary statistics (e.g., average *d*′ and accuracy) where in some cases *d*′ was low (e.g., Evans et al. 2016, Experiment 2, average *d*′ = .59 for the contralateral breast condition). It is possible that these results may have been driven by a few of the images. To fully test such a question, one needs to rule out any other explanation for the apparent lack of localisation in the key images (i.e., those in

which detection was accurate but localisation was not). These factors include response imprecision (coarse localisation information may be present) and correct guessing for the detection response (as outlined in the Howe and Webb (2014) paper reviewed above). Exploring these issues is a major theme of my work in Chapter 3.

Overall, a theme that runs through this thesis is the thorough examination of the claims that detection and localisation are dissociable at brief durations (e.g., Evans et al., 2013). I do this using natural scenes (Chapter 2) as these are an excellent model to use when investigating these processes, as well as medical images (Chapters 3 and 4), and using different paradigms: variants on brief visual search and attention cueing.

1.7 Overview of thesis

1.7.1 Chapter 2: The time course of rapid target detection and localisation

Visual search in natural scenes and medical images depend on the same underlying processes, making natural scenes an ideal model for radiologists' search in medical images. As reviewed above, previous studies have shown that a large amount of information can be extracted after viewing a natural scene briefly, such as scene categorisation (Greene & Oliva, 2009; Joubert, et al., 2007; Potter, 1976) and object detection (e.g., Thorpe, et al., 1996; VanRullen & Thorpe, 2001), but less is known about object localisation, which is important for goal-directed actions. This fast-visual processing has also been demonstrated in medical image interpretation (Evans, et al., 2013; Evans, et al., 2016; Kundel & Nodine, 1975; Nodine, et al., 1999). As previous studies in the medical domain have suggested no evidence for localisation of abnormalities at brief exposure durations even when detection is possible (Evans et al., 2013; Evans et al., 2016), the aim of the experiments reported in Chapter 2 was to test

whether a target could be localised in brief displays and to investigate its time course along with target detection. This chapter is not about natural scenes *per se*, but these are useful stimuli to explore this question, particularly as scene categorisation provides an independent verification that 'gist' has been extracted.

I presented natural scenes that varied on different sub-categories within the superordinate categories of natural and man-made. Using a two-alternate forced paradigm (2AFC) to reduce response bias, a target Gabor was superimposed (location randomised) and presented within either Scene 1 or Scene 2, which were presented for durations of 33-199ms. The observers were asked to detect the Gabor and then to localise it. In two experiments, I showed that a target could be localised when viewing a scene at very brief exposure durations (33ms). Experiment 2 was motivated from the findings from Experiment 1, which suggested that clutter mediated performance. It is well known that clutter impedes visual search during free viewing (Rosenholtz, et al., 2005), so we manipulated clutter using the computational definition of *enclosure* for natural scenes (Oliva & Torralba, 2001). The results showed target detection and localisation accuracy was higher for scenes that were open (e.g., coast) rather than closed (e.g., city centre). A plausible interpretation is for the open scenes, the Gabor may have provided a strong bottom-up signal, appearing salient. Overall the results presented in Chapter 2 demonstrate that a target can be detected and localised when viewing a scene briefly, and performance is mediated by visual clutter. Crucially, I showed localisation information could be extracted at the durations that have previously been documented to support detection (e.g., Thorpe et al., 1996; VanRullen & Thorpe, 2001).

1.7.2 Chapter 3: Finding cancer in mammograms: if you know it's there, do you know where?

The aim of the experiments reported in Chapter 3 is to investigate whether experienced radiologists can detect and localise a mass in a mammogram after seeing the image briefly. In Chapter 2, we found evidence of localisation information after viewing briefly presented natural scenes. Here, we investigated mammograms to explore the claims made by Evans et al. (2013). I adapted the task used in Evans et al. (2013) and presented radiologists with brief displays containing mammograms. In a pilot study, the radiologists saw images where 50% contained an easy-to-detect mass at 3 presentation durations (250ms, 500ms and 750ms, not backward masked). Participants were asked to decide whether there was an abnormality (detection) and then to locate the mass on a blank outline of the mammogram (localisation). I found that detection and localisation performance was above that expected by chance, providing a firm basis for a further study using a similar paradigm. In the main experiment, I built upon these findings to explore how breast density moderates the type of information extracted in a brief display. I increased task difficulty by including images with less obvious masses due to higher breast density (50% high density) to more closely replicate the paradigm of Evans et al. (2013) and presented mammograms at 2 durations (250 and 1000ms, backward masked). I found evidence for detection and localisation, even at our briefest duration (250ms). This is consistent with previous findings that those with experience can extract a considerable amount of information to support detection in the first quarter of a second when viewing an image, but also demonstrates that localisation information can be extracted at these durations.

Although localisation performance was above chance in my studies, there was a proportion of trials on which detection appeared to occur in the absence of localisation

(as in Evans et al., 2013 and Evans et al., 2016). This therefore offered the potential to fully explore the possibility of a dissociation between detection and localisation. An analysis of individual location error trials is crucial for the question of dissociation; it cannot be answered using simple summary statistics and frequentist analyses. I categorised localisation error types and discovered a number of factors that led to the underestimation of localisation including stimulus variability, response imprecision and participant guesses. This study also showed that higher breast density reduced performance. Breast density is a related (not wholly equivalent) construct to clutter and here we replicated what is known from the visual search literature that clutter reduces performance (e.g., Rosenholtz, et al., 2005), extending these findings to medical images. These findings are also consistent with the findings in Chapter 2 (using natural scenes as stimuli), where I showed that performance is mediated by clutter in a scene. Broadly speaking, these results differ from the Evans et al. (2013) account and provide support for the Kundel and Nodine (1975) model of visual search.

1.7.3 Chapter 4: The influence of prior expectation and expertise on attentional cueing in medical images

Another method for indirectly measuring whether information about location of a mass is extracted is to look for the impact of a mass on a subsequent detection task. In this chapter, I used an attention cueing task where medical images served as primes. This allowed me to further explore localisation, but also examine some important questions of how prior knowledge and expertise might influence the allocation of attention in medical images.

Radiologists make critical decisions based on searching and interpreting medical images and the probability of abnormalities varies across anatomical regions in the

body. The allocation of spatial attention may be influenced in a top-down manner, by priors that may set a search strategy or attentional bias. The aim of the experiments reported in Chapter 4 is to investigate the extent to which expertise guides attention based on prior experience with the prevalence of cancer, using a cueing paradigm where a chest radiograph (with or without a suspicious nodule) formed a prime. This experiment also addressed whether localisation information was available as indexed by attentional capture and if expertise boosted the salience of subtle nodules that do not affect attentional allocation in naïve observers. In Experiment 1, with naïve observers, an artificially boosted nodule in the prime radiograph guided attention, validating the task. In Experiment 2, radiologists viewing real nodules did not show the same effect, nor did they show any attentional guidance from cancer prevalence. However, more experienced radiologists seemed to be more sensitive to the subtle nodules than less experienced radiologists, suggesting that expertise might boost nodule salience.

1.7.4 Chapter 5: General discussion

In Chapter 5, I summarise and draw together the major findings from these experiments: target localisation is possible along with detection in both natural scenes and medical images and decisions can be influenced by experience. In this chapter I discuss the theoretical implications of these findings, and how they fit within the current literature specifically in the context of the two major models of radiologist visual search. I discuss implications in a broader scientific and clinical context, the challenges and limitations of my research and future directions, before drawing it all together with general conclusions.

1.8 References

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2013). Self-induced attentional blink a cause of errors in multiple-target search. *Psychological science*, *24*(12), 2569-2574.

Al Mousa, D. A., Ryan, E. A., Mello-Thoms, C., & Brennan, P. C. (2014). What effect does mammographic breast density have on lesion detection in digital mammography? *Clinical radiology*, *69*(4), 333-341.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367-10371.

Anderson, B. A., & Yantis, S. (2013). Persistence of value-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 6.

Becker, S. I. (2008). Can inter-trial effects of features and dimensions be explained by a single theory? *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1417.

Becker, S. I., Folk, C. L., & Remington, R. W. (2010). The role of relational information in contingent capture. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1460.

Berbaum, K. S., Franken Jr, E. A., Dorfman, D. D., Rooholamini, S. A., Kathol, M. H., Barloon, T. J., Behlke, F.M., Sato, Y., Lu, C.H., El-khoury, G.Y., & Flickinger, F. W. (1990). Satisfaction of search in diagnostic radiology. *Investigative radiology*, *25*(2), 133-140.

Berlin, L. (2005). Errors of omission. *American Journal of Roentgenology*, *185*(6), 1416-1421.

Boyd, N. F, Martin, L. J, Bronskill, M, Yaffe, M. J, Duric, N & Minkin S. (2010). Breast tissue composition and susceptibility to breast cancer. *J Natl Cancer Inst, 102*, 1224-37.

Broadbent, D.E. (1958). *Perception and Communication*. Oxford: Pergamon.

Charness, A. N. (2014). The Role of Practice and Coaching in Entrepreneurial Skill Domains: An International Comparison of LifeSpanChess Skill. *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive psychology*, *36*(1), 28-71.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559-564.

Della Libera, C., & Chelazzi, L. (2006). Visual selective attention and the effects of monetary rewards. *Psychological science*, *17*(3), 222-227.

Drew, T., Evans, K., Võ, M. L. H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics*, *33*(1), 263-274.

Drew, T., Võ, M. L. H., Olwal, A., Jacobson, F., Seltzer, S. E., & Wolfe, J. M. (2013). Scanners and drillers: Characterizing expert visual search through volumetric images. *Journal of vision*, *13*(10), 3-3.

Drew, T., Võ, M. L. H., & Wolfe, J. M. (2013). The invisible gorilla strikes again sustained inattentional blindness in expert observers. *Psychological science*, 0956797613479386.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological review*, *96*(3), 433.

Evans, K. K., Birdwell, R. L., & Wolfe, J. M. (2013). If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening. *PLoS One*, *8*(5), e64366.

Evans, K. K., Georgian-Smith, D., Tambouret, R., Birdwell, R.L., & Wolfe, J.M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin Review.* doi: 10.3758/s13423-013-0459-3

Evans, K. K., Haygood, T. M., Cooper, J., Culpan, A.-M., & Wolfe, J. M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences*, *113*(37), 10292–10297. http://doi.org/10.1073/pnas.1606187113.

Fernandez-Duque, D., & Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition, 7,* 323–344.

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human perception and performance*, *24*(3), 847.

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology-Human Perception and Performance*, *20*(2), 317-329.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137-176.

Handford, M. (1987). *Where's Waldo?* Boston: Little, Brown & Co.

Hidalgo-Sotelo, B., Oliva, A., & Torralba, A. (2005, June). Human learning of contextual priors for object search: where does the time go?. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on* (pp. 86-86). IEEE.

Houghton, J. P., Smoller, B. R., Leonard, N., Stevenson, M. R., & Dornan, T. (2015). Diagnostic performance on briefly presented digital pathology images. *Journal of Pathology Informatics*, *6*, 56. http://doi.org/10.4103/2153-3539.168517

Howe, P.D.L. & Webb, M.E. (2014). Detecting unidentified changes. *PloS ONE, 9*(1), e84490.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. *Attention and performance IX*, *9*, 187-203.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision research*, *47*(26), 3286-3297.

Kristjánsson, A. (2006). Simultaneous priming along multiple feature dimensions in a visual search task. *Vision Research*, *46*(16), 2554-2570.

Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiol. 4***,** 219–227.

Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of vision*, *14*(3), 14-14.

Kundel, H. L., & La Follette Jr, P. S. (1972). Visual search patterns and experience with radiological images. *Radiology*, *103*(3), 523-528.

Kundel, H. L., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology*, *116*(3), 527-532.

Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative radiology*, *13*(3), 175-181.

Kundel, H. L., Nodine, C. F., Krupinski, E. A., & Mello-Thoms, C. (2008). Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic radiology*, *15*(7), 881-886.

Lamy, D., Zivony, A., & Yashar, A. (2011). The role of search difficulty in inter-trial feature priming. *Vision Research*, *51*(19), 2099-2109.

Lopez, N., Previc, F. H., Fischer, J., Heitz, R. P., & Engle, R. W. (2012). Effects of sleep deprivation on cognitive performance by United States Air Force pilots. *Journal of Applied Research in Memory and Cognition*, *1*(1), 27-33.

Mack, A., & Rock, I. (1998). Inattentional Blindness: Perception without attention. Cambridge, MA: MIT Press.

Mello-Thoms, C. (2009). The "Holistic Grail": Possible implications of an initial mistake in the reading of digital mammograms. *Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment (Proceedings of SPIE, Vol. 72630)* doi: 10.1117/12.813778

Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., Stalder, J., & Maitz, G. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Academic radiology*, *12*(7), 830-840.

Mitroff, S. R., & Biggs, A. T. (2014). The ultra-rare-item effect visual search for exceedingly rare items is highly susceptible to error. *Psychological Science*, *25*(1), 284-289.

Mitroff, S. R., Simons, D. J., & Franconeri, S. L. (2002). The siren song of implicit change detection. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 798.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Nodine, C. F., & Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Academic radiology*, *5*(9), 603-612.

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C., & Conant, E. F. (1999). How experience and training influence mammography expertise. *Academic radiology*, *6*(10), 575-585.

Nodine, C.F., & Mello-Thoms, C. (2010). The role of expertise in radiologic image interpretation (p 139-156). In *The handbook of medical image perception and techniques*. Samei, & E. A. Krupinski (Eds.). Cambridge: Cambridge University Press.

Oliva, A. (2005). Gist of the scene. *Neurobiology of Attention*, *696*(64), 251–258.

Oliva, A. & Torralba, A. (2001). Publication utilizing eight semantically organized categories: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. From http://cvcl.mit.edu/database.htm

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3-25.

Potter, M.C. (1976). Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem. 2*(5), 509–522.

Rawashdeh, M. A., Lee, W. B., Bourne, R. M., Ryan, E. A., Pietrzyk, M. W., Reed, W.M., Heard, R.C., Black, D.A. & Brennan, P.C. (2013). Markers of good performance in mammography depend on number of Annual readings. *Radiology, 269*, 61 – 67.

Ray, S., Chen, L., Keller, B. M., Chen, J., Conant, E. F., & Kontos, D. (2016). Association between Breast Parenchymal Complexity and False-Positive Recall From Digital Mammography Versus Breast Tomosynthesis: Preliminary Investigation in the ACRIN PA 4006 Trial. *Academic radiology*, *23*(8), 977-986.

Rensink, R. A. (2004). Visual sensing without seeing. *Psychological Science*, *15*(1), 27-32.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368–373.

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005, April). Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 761-770). ACM.

Samei, E., & Krupinski, E. A. (2010). In *The handbook of medical image perception and techniques*. Samei, & E. A. Krupinski (Eds.). Cambridge: Cambridge University Press.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059-1074.

Smilek, D., Eastwood, J. D., & Merikle, P. M. (2000). Does unattended information facilitate change detection? *Journal of Experimental Psychology: Human Perception and Performance, 26,* 480–487.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582*),* 520 – 522.

Torralba, A., Murphy, K. P., & Freeman, W. T. (2010). Using the forest to see the trees: exploiting context for visual object detection and localization. *Communications of the ACM*, *53*(3), 107-114.

Treisman, A.M. (1964). Verbal cues, language, and meaning in selective attention. *American Journal of Psychology, 77*, 242-248.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Tuddenham, W. J. (1962). Visual Search, Image Organization, and Reader Error in Roentgen Diagnosis: Studies of the Psychophysiology of Roentgen Image Perception Memorial Fund Lecture 1. *Radiology*, *78*(5), 694-704.

VanRullen, R., & Thorpe, S.J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artefactual objects. *Perception, 30*(6), 655-688.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.

Wolfe, J. M. (2001). Guided Search 4.0: A guided search model that does not require memory for rejected distractors. *Journal of Vision*, *1*(3), 349-349.

Wolfe, J. M. (2012). When do I quit? The search termination problem in visual search. In *The influence of attention, learning, and motivation on visual search* (pp. 183-208). Springer New York.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, *15*(3), 419.

Wolfe, J. M., Evans, K. K., Drew, T., Aizenman, A., & Josephs, E. (2016). How do radiologists use the human search engine? *Radiation protection dosimetry*, *169*(1-4), 24-31.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*, 0058.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare targets are often missed in visual search. *Nature*, *435*(7041), 439.

Wolfe, J.M., Võ, M.L., Evans, K.K., & Greene, M.R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Science, 15*(2), 77-84.

Chapter 2: The time course of rapid target detection and localisation

The time course of rapid target detection and localisation

Ann. J. Carrigan[1,2,3], Susan. G. Wardle[1,2], Anina. N. Rich[1,2,3].

[1]*Perception in Action Research Centre & Department of Cognitive Science, Macquarie University, Australia*

[2]*ARC Centre of Excellence in Cognition & Its Disorders, Macquarie University, Australia*

[3]*Centre for Elite Performance, Expertise, and Training, Macquarie University, Australia*

The manuscript that describes this study was co-authored by Carrigan, A.J., Wardle, S.G., & Rich, A.N. (under review following revision). The time course of rapid detection and localisation of a target in natural scenes. *Journal of Vision.*

Corresponding author: ann.carrigan@mq.edu.au

Abstract

The human visual system is capable of processing an enormous amount of detail in a short space of time. Scene categorisation and target detection are possible after only brief exposures. Although rapid target detection has been explored extensively, less is known about target localisation. In addition to processing *what* objects are present in a scene, information about *where* each object is located is important for guiding goal-directed action. We measured the time course of target detection and localisation in natural scenes. Participants viewed scenes across four presentation durations (33-199ms) taken from two superordinate categories (natural and man-made), each containing exemplars from four basic scene categories. In a 2AFC task, observers were asked to detect a Gabor inserted in one of the two scenes. This was followed by one of two different localisation tasks. Participants were either asked to discriminate whether the target was on the left or the right side of the display, or to click on the exact location where they saw the target. The results demonstrate that a target can be detected and localised at exposures as brief as 33ms, with performance improving with increasing exposure duration. Experiment 2 showed that closed scenes which typically contained higher levels of visual clutter than open scenes reduced the accuracy of detection and localisation. The results demonstrate that a target can be localised when viewing a scene at very brief exposure durations, and performance is mediated by visual clutter. This has real-world implications for visual search tasks such as medical screening.

Keywords: fast visual processing, visual clutter, target detection, target localisation

Vision is fast: as soon as our eyes open we get an impression that we can see everything around us. Early findings suggested that the basic meaning of natural scenes (e.g., classification as 'outdoor' versus 'indoor' scenes) can be extracted after an exposure of only 100ms (Potter, 1976; Potter & Faulconer, 1975). Further studies using backward masking to precisely control display duration showed that observers are above chance at categorising scenes at the superordinate (e.g., natural vs. man-made) and basic (e.g., coast vs. city) levels after exposure durations as short as 20ms (Greene & Oliva, 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007). In addition, when primed with an object category (e.g., animal or truck) these objects can be accurately detected when observers are shown scenes for only 20 -25ms, albeit with no backward mask (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). Moving from the laboratory to real-world tasks, at slightly longer durations, experts can process medical images to a remarkable degree. Kundel & Nodine (1975) showed that when presented with a chest radiograph for 200ms, radiologists could detect an abnormality with 70% accuracy. Thus, there is evidence that considerable information is extracted in the initial glance at an image or environment.

Categorisation of natural scenes at these brief durations (Oliva, 2005; Potter, 1976), and, more recently, detection of abnormalities in medical images (Evans, Georgian-Smith, Tambouret, Birdwell, & Wolfe, 2013; Evans, Haygood, Cooper, Culpan & Wolfe, 2016), has been argued to occur based on a global signal or 'gist'. Although it seems intuitive that information about *where* an item is should accompany detection of its presence, the global nature of 'gist' implies that specific location information is unlikely to be available. Indeed, previous studies on change detection in faces (Howe & Webb, 2014) and abnormality detection within medical images (Evans, et al., 2013; Evans, et al., 2016) have suggested that target detection is possible without localisation under some

circumstances. Howe and Webb (2014) showed observers a photograph of a face for 1.5 seconds, followed by a 1 second blank, and then another version of the same photograph with a single changed feature (e.g., removal of glasses). Observers were asked to indicate if a change had occurred and if so, to select the change from a list of nine possible options. The results showed that observers could sometimes detect that a change had occurred without identifying the specific change, even when taking into account potential correct guesses. The authors suggest that the apparent lack of information about the identity of the change might reflect low precision in the location. In contrast, other change detection studies have found that detection of a change is accompanied by knowledge of the change location, and that this performance is driven by feature salience (Mitroff & Simons, 2002).

There has been considerable interest in detection and localisation within medical images. Eye-gaze data has shown that abnormalities in chest radiographs and mammograms can be detected prior to fixation on the lesion, with subsequent search leading to localisation within one to two seconds of image onset (Kundel, Nodine, Conant & Weinstein, 2007; Mello-Thoms, et al., 2005). More recent studies have suggested that, at relatively brief durations, expert radiologists may be able to detect an abnormality but not locate it (Evans et al., 2013; Evans et al., 2016). Evans et al. (2013) compared the performance of novices and radiologists on the detection and localisation of abnormalities. The stimuli were mammograms containing subtle masses and architectural distortions that varied in size (10 to 48mm). They presented bilateral images (left and right breasts) at durations from 250ms to 2000ms and showed that radiologists (but not novices) could detect an abnormality within a mammogram above chance. Following detection, the radiologists viewed a blank outline of the mammogram and were asked to localise the abnormality with the computer mouse. Chance was determined by calculating

the average (across images) percentage of overall tissue area lying within a predetermined region of abnormality. At a display duration of 500ms, abnormalities could be detected by radiologists above chance but localisation performance was not statistically different from chance. Using a similar design and stimuli, Evans et al (2016) replicated these findings: at 500ms detection was greater than chance, with localisation accuracy at 21%, not significantly different from chance performance. This raises the possibility that in some cases, at brief durations, target detection may occur in the absence of information sufficient to locate the target.

The aim of the present study is to investigate the time course of target detection and localisation in brief displays. As previous studies in the medical domain have found no evidence for localisation of abnormalities at brief exposure durations even when detection is possible (Evans et al., 2013; Evans et al., 2016), we aimed to evaluate whether any localisation information is accessible at very short durations. To adapt the paradigm used in medical imaging to a broader visual context, we used natural scenes as these provide an independent measure where some meaning-level information can be processed. This category of stimuli is one we are all experts at viewing and interpreting, and previous research has already demonstrated we can detect targets (e.g., animals) within scenes at brief durations (Thorpe, et al., 1996; VanRullen & Thorpe, 2001). We can also verify that 'gist' level information has been extracted at our selected durations by using a scene classification task. We embedded a target Gabor within the scenes rather than using an identifiable nameable object within the scene. This was to avoid invoking any semantic congruency, as the Gabor is equally irrelevant in the different scenes (see Davenport & Potter, 2004). This study is not about natural scenes *per se*, using this paradigm allows precise experimental control over factors such as target location, size

and luminance contrast, which is not possible with real world stimuli such as abnormalities in medical images.

To validate the durations we used, in Experiment 1A we confirm that scene categorisation (natural vs. man-made) is possible for the background scenes at the shortest experimental duration. In Experiment 1B, we then compare detection and localisation performance for a Gabor target embedded in a range of natural scenes at brief exposure durations between 33ms and 199ms. Based on the results of Evans et al. (2013) and Evans et al. (2016) with medical images, we test for evidence of information sufficient to support detection and localisation at each duration. To test the idea that location information might be present but less precise at brief durations (Howe & Webb, 2014), we included a Left (L) vs Right (R) localisation task as well as the 'click on the location with the mouse' version used by Evans et al. (2013) and Evans et al. (2016). Finally, in Experiment 2 we investigate the influence of visual clutter in natural scenes on target detection and localisation at brief durations.

Experiment 1

Experiment 1A was a scene categorisation task (natural vs. man-made) designed to verify that the overall 'gist' of the background scenes could be extracted at the shortest experimental duration (33ms) used in our paradigm. The aim of Experiment 1B was to test whether durations between 33 and 199ms resulted in sufficient processing to support detection and localisation of a Gabor target embedded in the natural scenes. Using a two-alternative forced paradigm (2AFC), natural scenes were presented at one of four durations (33 – 199ms) on each trial, with a Gabor target randomly located within one of the two scenes. The participants were asked to report which scene contained the target and then where it was located within the scene.

Method

All measures and conditions are reported.

Participants

Thirty participants (22 females, age range 19 - 55 years, mean age [*M*] = 31.47 years, standard deviation [*SD*] = 8.81) were recruited from Macquarie University. All participants gave informed consent, reported normal or corrected-to-normal vision, and were financially reimbursed for their time. The study was approved by the Macquarie University Human Research Ethics Committee (Medical Sciences). The data for two observers were excluded due to technical issues, leaving 28 datasets for analysis.

Stimuli and Apparatus

Natural scene stimuli were classified as scene categories defined by Oliva and Torralba (2001), which are available at http://cvcl.mit.edu/database.htm (see Fig. 1). One hundred and sixty photographic images of natural scenes comprising two superordinate categories (natural and man-made) were selected from an internet search using Google Images. The natural and man-made categories comprised four basic-level categories (20 images in each): coast, mountain, open country and forest for the natural category, and tall building, highway, city centre and street for the man-made category. The images were converted to greyscale and downsized to subtend 23° x 15° of visual angle.

Figure 1. Example scenes for the superordinate categories and four basic categories within each.

The target was a Gabor patch with the following parameters: orientation 45°; spatial frequency 0.5° cycles/degree; diameter 3.8°; Michelson contrast 0.2. The target image appeared in a different random location within the scene boundaries and was present on all trials in either Scene 1 or Scene 2.

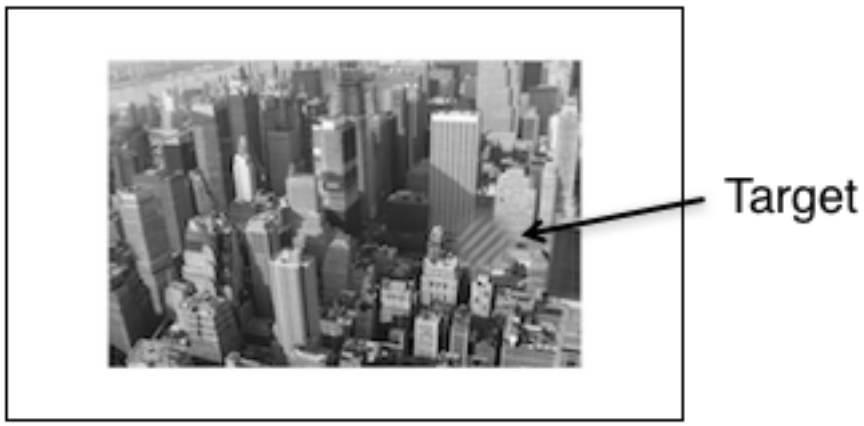


Figure 2. Example of a city (manmade) scene with the embedded target Gabor (diameter = 3.8°).

The participants sat at a viewing distance of approximately 70cm in a dimly lit, windowless laboratory at Macquarie University, Sydney. Stimuli were presented with MATLAB 8.2. using PsychToolbox 3 (Brainard, 1997; Pelli, 1997) and were displayed on a 27-in Samsung SyncMaster SA950 LCD monitor (1920 x 1080, 120 Hz).

Procedure

Experiment 1A: Scene Categorisation Task

This task was to verify that scene categorisation (natural vs. man-made) is possible for the background scenes at the shortest experimental duration. We used a single factor (superordinate category: natural, manmade) within-subjects design. Each trial began with a fixation point for 498ms, followed by a scene from one of the superordinate categories displayed in the centre of the screen for 33ms. This was followed by a backward 1/f noise mask for 249ms. Participants categorised the scene by its superordinate category ("manmade" vs. "natural") with a key press as accurately and quickly as possible (See Fig. 3). Participants were given ten practice trials at a longer scene presentation duration of 398ms to familiarise them with the task before completing the experimental trials.

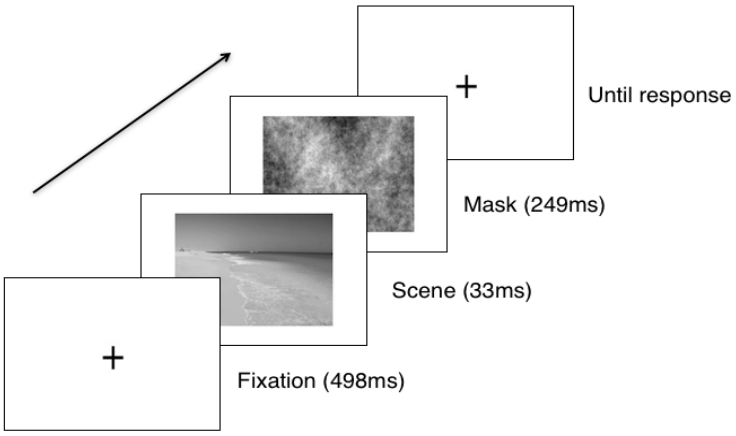


Figure 3. Example of a trial for the superordinate scene categorisation task in Experiment 1A. The participant responded whether the image was a manmade or natural scene.

Experiment 1B: Target detection and localisation task.

For the main experimental task, we used a 2 (location task: exact click or L vs. R) x 4 (duration: 33, 58, 116, 199ms) within-subjects design. Initially the participants were shown a picture of the target to familiarise them with a Gabor, and given eight practice trials (2 per duration) with feedback. Each trial began with a fixation point for 498ms, followed by Scene 1 [33 – 199ms, constant within a block] followed by a backward 1/*f* noise mask for 249ms and then Scene 2 [same duration as Scene 1] followed by a 1/*f* noise mask [249ms]. Observers made a 2AFC decision with a key press regarding whether the target was present in Scene 1 or Scene 2. Following this detection response, they were presented with a blank screen and asked one of two localisation questions (in separate blocks; order counterbalanced across participants). In one block of trials, they were asked to click on the exact location of the target on the blank screen using the mouse. In the other localisation task, they were instead asked whether the target appeared

on the left or the right side of the screen and responded using a key press. This localisation task required a coarser judgment of the target's location in order to answer correctly, compared to the more difficult exact click task. The response keys for the left/right localisation task (L/R) were the same as the keys used for the detection task (Keys ‘z’ and ‘m’; see Fig. 4). On each trial, both scenes were selected from the same superordinate category (e.g., natural or man-made) but the basic category was random (e.g., both could be from the same category or from different categories within the superordinate category). Fifty percent of trials had the Gabor in Scene 1 and 50% in Scene 2, randomly interleaved within a block. Target location was randomised, with the restriction that it was not clipped by the screen edge and that it appeared in the left half for 50% of trials, and in the right for the other 50% of trials. Duration order was blocked and counterbalanced across participants. The participants performed 160 experimental trials for each localisation task. The experiment was self-paced and the participants initiated each trial with a key press. The observers saw the same images in each task, but in a different randomised order (80 natural and 80 man-made scenes in each version of the task), giving a total of 320 trials across the experiment. They were instructed to respond as accurately as possible and there was a minimum 15 second rest period every 40 trials. Participants were not provided with any feedback during the experimental tasks (see Fig. 4).

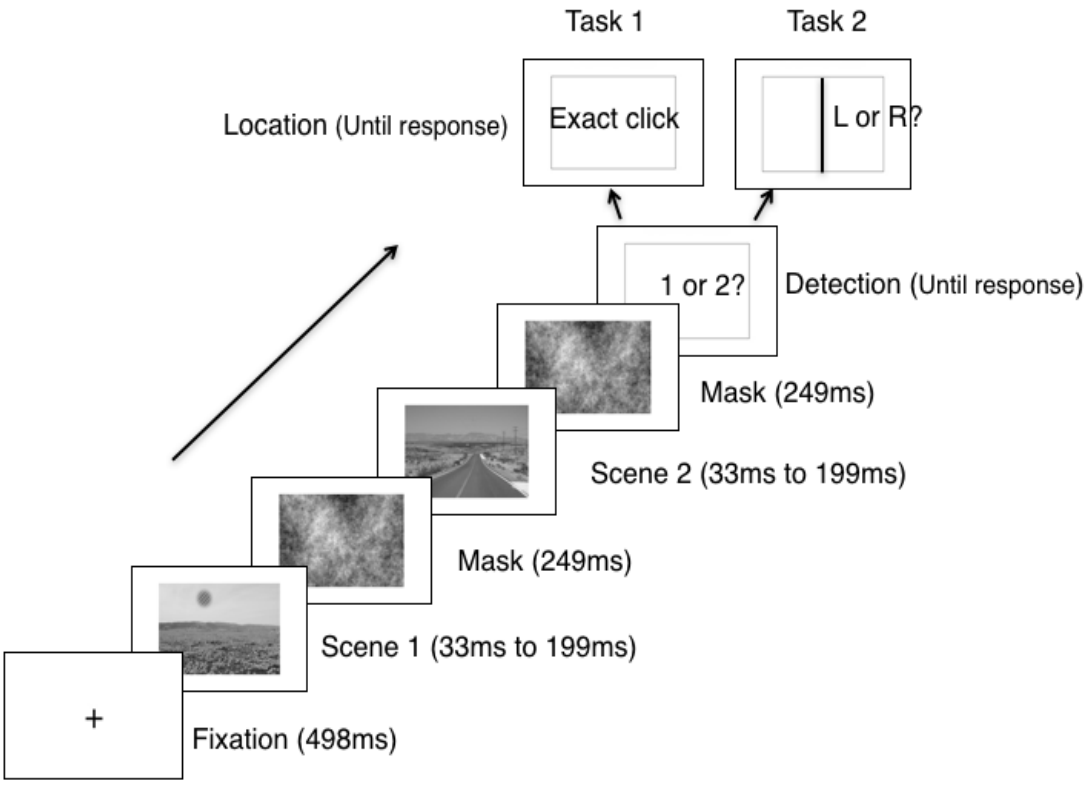


Figure 4: Example of a trial for the target detection and localisation tasks, with the Gabor target in Scene 1. Note the order of the task blocks was counterbalanced across participants and the target was as equally likely to appear in Scene 1 as Scene 2.

Results and Discussion

Experiment 1A: Scene categorisation

We first verified that sufficient information about scene categories (natural vs man-made) could be extracted at our shortest duration (33ms), as has been demonstrated by others (Greene & Oliva, 2009; Joubert et al., 2007). We used a measure of sensitivity, *d′,* as our dependent measure. Mean *d′* for the categorisation tasks was 2.29 (*SD* = .56, *range* = 1.36 – 3.65). A single sample t-test on *d′* relative to chance (*d′* = 0) demonstrated performance was better than chance in categorising the scenes as manmade vs. natural at

the shortest experimental duration, $t(27) = 21.8$, $p < .0001$. This replicates previous findings that sufficient visual information to categorise scenes is available from 33ms presentations.

Experiment 1B

Detection performance

Figure 5 shows detection performance for the Gabor target at exposure durations between 33ms and 199ms. We calculated *d* prime as a function of target presence in Scene 1 or Scene 2. A two-way repeated measures ANOVA on d' with the factors of Localisation Task (exact click, L vs. R) x Duration (33, 58, 116, 199) revealed no main effect of Localisation Task, $F(1, 27) = .38$, $p = .541$, a significant main effect for Duration, $F(1.1, 29.79) = 16.25$, $p < .0001$, $\eta^2_p = .38$ (Greenhouse-Geisser corrected), and no significant Localisation Task x Duration interaction, $F(1.1, 29.71) = 1.29$, $p = .27$, $\eta^2_p = .046$ (Greenhouse-Geisser corrected). The detection task was identical for the two location tasks and detection was performed prior to the localisation task. It is therefore not surprising that we see only an effect of improved performance as duration increased. Our primary question for target detection is whether at each duration there is sufficient information to support detection. We therefore collapsed the detection data across localisation task and evaluated detection performance using single sample t-tests at each duration relative to a chance level of $d' = 0$. Figure 5 shows the data for detection performance (d prime) collapsed across localisation tasks. To maintain an overall Type I error rate of .05, a Bonferroni correction was used (test-wise alpha was set at $p = .0125$). Detection performance was significantly above chance at each exposure duration [33ms, $t(27) = 3.98$, $p < .0001$; 58ms, $t(27) = 16.04$, $p < .0001$; 116ms, $t(27) = 27.51$, $p < .0001$; 199ms, $t(27) = 34.23$, $p < .0001$].

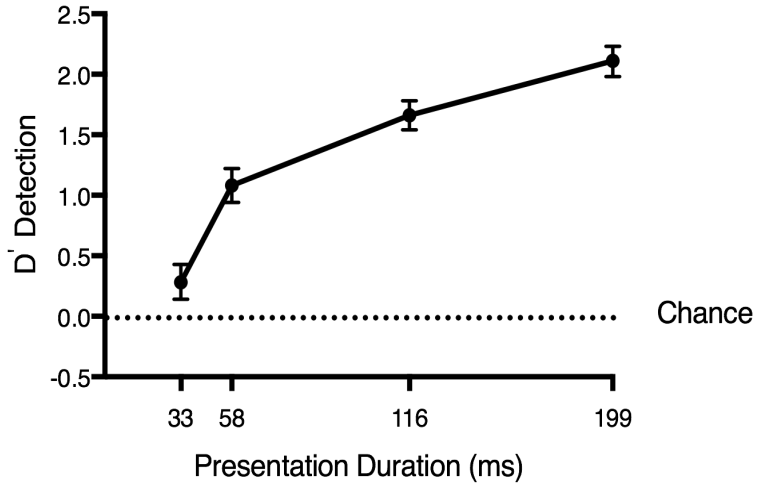


Figure 5. Detection performance collapsed across location tasks. The dashed line represents chance. The error bars represent 95% confidence intervals.

Localisation performance

The results above show that the target Gabor could be detected in complex natural scenes with presentations even as brief as 33ms. These results are consistent with the previous literature that has shown accurate object detection within scenes at exposure durations between 20-25ms (Thorpe et al., 1996; VanRullen & Thorpe, 2001). Next, we investigated whether the target can also be located at these very brief presentation durations. Our dependent variable was percentage localisation correct. We analysed the total percentage of localisation correct across all trials, splitting the analysis by the two localisation tasks: L vs. R (coarse localisation) and exact click (fine localisation), which were presented in separate blocks.

Localisation in the coarse L vs. R task: This localisation task was a 2AFC: left or right. Thus, chance is 50%. Figure 6 shows a clear pattern of increasing localisation performance with increasing duration for both localisation tasks. Our key question relates

to whether there is sufficient information at each duration to support localisation. We therefore used single sample t-tests on percent correct localisation responses (Bonferroni corrected, test-wise alpha was set at $p = .0125$). For the L vs. R task (chance = 50%), this showed that performance was above chance for all durations [Fig 6, black line; 33ms, $t(27) = 5.07$, $p < .0001$; 58ms, $t(27) = 11.10$, $p < .0001$; 116ms, $t(27) = 27.49$, $p < .0001$; 199ms, $t(27) = 15.87$, $p < .0001$].

Localisation in the exact click localisation task: This localisation task was a precise mouse click on the target. We calculated chance based on the number of possible non-overlapping locations of the target Gabor within the image (chance = 16.67%). To allow for some imprecision in reporting the remembered target location, we defined a region of acceptance (ROA) for scoring a mouse click as correct localisation of twice the Gabor diameter: 7.6° centred on the Gabor location. Our dependent variable was percentage localisation correct. Again, our key question relates to whether there is sufficient information at each duration to support localisation. Single sample t-tests were conducted on percent correct localisation responses (Bonferroni corrected, test-wise alpha was set at $p = .0125$). For the exact click task (chance = 16.67%), this showed that performance was above chance for all durations [Fig 6, blue line; 33ms, $t(27) = 6.3$, $p < .0001$; 58ms, $t(27) = 14.78$, $p < .0001$; 116ms, $t(27) = 26.75$, $p < .0001$; 199ms, $t(27) = 51.21$, $p < .0001$].

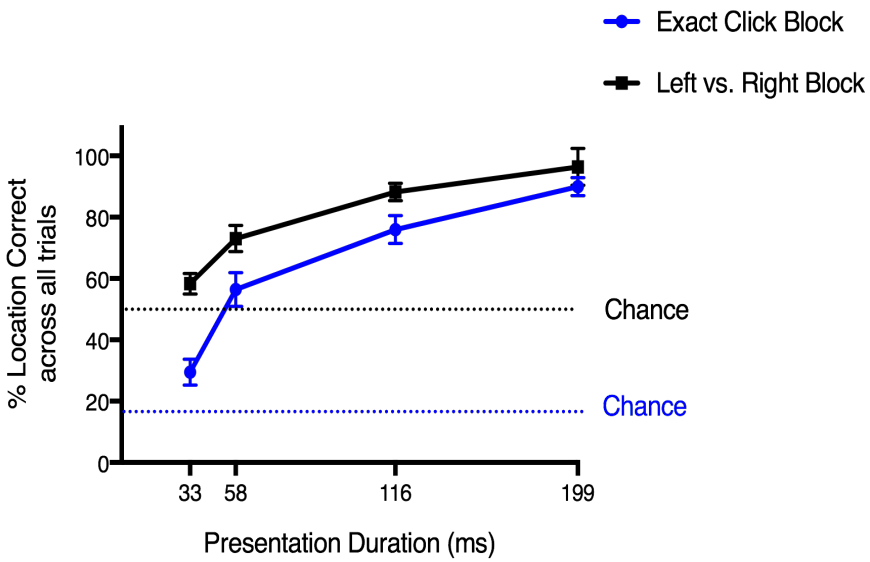


Figure 6. Accuracy on the two localisation tasks. The dashed line represents chance (black: Left vs. Right; blue: exact click). The error bars represent 95% confidence intervals. Note: Due to the different features of each task is it not possible to directly compare performance.

The localisation results show that the participants could accurately localise a Gabor target on some trials for presentation durations as brief as 33ms. Specifically, participants performed significantly better than chance for all durations, 33-199ms, even for the precise localisation task (exact click). In Figure 7 and the corresponding analysis, all localisation trials are included regardless of whether detection was correct. Figure 7 represents the relative proportion of trials across all four durations when detection and localisation were both correct (dark blue bar), when detection was correct and localisation was incorrect (light blue bar), when localisation was correct and detection was incorrect (dark grey bar) and when detection and localisation were both incorrect (light grey bar).

Each experimental trial is represented once in the graph. The proportion of trials on which both detection and localisation are correct clearly increases as a function of duration, as one would expect. At the shortest durations, accuracy on detection and localisation appears lower in absolute performance for the more precise localisation task (exact click) compared with the coarse localisation task (L vs. R), however note that differences in chance baseline between the two localisation tasks (50% for left/right, 16.67% for exact click) limits a direct comparison of between-task performance. Figure 7a for the L vs. R task shows that at 33ms duration, the observers actually got ~20% of trials correct on *localisation* when they got the corresponding detection task incorrect. This unexpected finding, however, is likely to be due to a keyboard assignment issue. The keys for the detection response and localisation were the same (e.g., 'z' = Scene 1 and Left; 'm' = Scene 2 and Right). This seems to have caused a 'response conflict' effect where participants are giving an initial response incorrectly based on localisation rather than which scene contained the target. We therefore do not interpret these localisation-without-detection trials for the left/right task further.

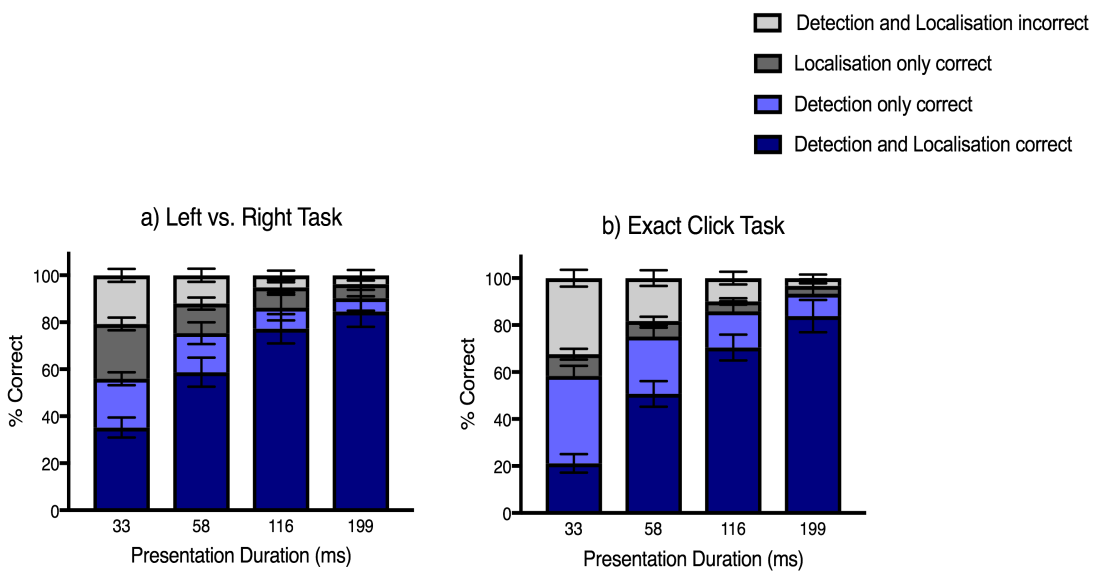


Figure 7. Percentage of trials across all four durations grouped as a function of response profile: trials on which detection and localisation were both correct (dark blue bar), trials on which detection was correct and localisation was incorrect (light blue bar), trials on which localisation was correct and detection was incorrect (dark grey bar) and trials on which detection and localisation were both incorrect (light grey bar) for (a) Left vs. right task and (b) Exact click task. The error bars represent 95% confidence intervals.

Returning to the summary statistics, the overall finding that observers are greater than chance on both detection and localisation show that even at brief durations, a target embedded in a natural scene can often be spatially localised as well as detected. In the medical imaging domain, Evans et al. (2013) and Evans et al. (2016) reported detection in the absence of localisation. A critical difference between our stimuli and that of Evans et al. (2013) and Evans et al. (2016) is that we used an artificial target Gabor embedded in

natural scenes, whereas the targets in Evans et al. (2013) and Evans et al. (2016) were abnormalities in medical images, which are often subtle and less visually distinct from the healthy tissue background. Therefore, localisation of targets at brief presentations may depend on how salient the target is within a particular scene, or on features of particular types of scenes such as visual clutter. We performed an exploratory post-hoc scene analysis on the effect of visual clutter on localisation performance within our diverse natural scene image set.

As the definition of 'visual clutter' is not straightforward in complex natural scenes, for our post-hoc analysis we operationalised clutter based on the related concept of *enclosure.* Within the computational literature, the global properties or distribution of basic features of a scene such as level of *enclosure* have been described using the Spatial Envelope Model (*SEM*; Oliva & Torralba, 2001). A scene that has a *closed* spatial envelope is composed of many visual characteristics (e.g., forest, city and mountain). In contrast, an *open* scene appears vast and clutter free with minimal visual items (e.g., coast, highway and open country). Enclosure is related to visual clutter within a scene, which can lead to lower accuracy in decision making tasks (Oliva, Mack, Shrestha & Peeper, 2004; Bravo & Farid, 2007; Rosenholtz, Li & Nakano, 2007).

If localisation is affected by visual clutter, there should be a difference between localisation performance on correct-detection trials in open versus closed scenes. We divided the scenes into categories according to the level of enclosure (open and closed) where open represents low clutter and closed represents high clutter and plotted the data separately (Figure 8). We conducted a two-way repeated measures ANOVA on the two localisation tasks separately with the factors of Scene (open, closed) and Duration (33, 58, 116, 199).

For the L vs. R task, there was a significant main effect of Scene (Open vs. Closed), $F(1,27) = 70.16$, $p < .0001$, $\eta^2_p = .72$, a significant main effect of Duration, $F(2.63, 71.11) = 176.13$, $p < .0001$, $\eta^2_p = .87$, (Huynh-Feldt corrected), and a significant Scene by Duration interaction, $F(2.21, 59.58) = 5.99$, $p = .001$, $\eta^2_p = .18$ (Greenhouse-Geisser corrected). Similarly, for the Exact click task, there was a significant main effect of Scene (Open vs. Closed), $F(1,27) = 87.49$, $p < .0001$, $\eta^2_p = .76$, a significant main effect of Duration, $F(3,81) = 259.69$, $p < .0001$, $\eta^2_p = .91$ and a significant Scene by Duration interaction, $F(3, 81) = 7.96$, $p < .0001$, $\eta^2_p = .23$. The interactions suggest that clutter does influence the degree to which location information is available. Experiment 2 was designed to follow up this initial analysis by experimentally manipulating the degree of visual clutter in natural scenes to systematically examine its effects on target detection and localisation at brief durations.

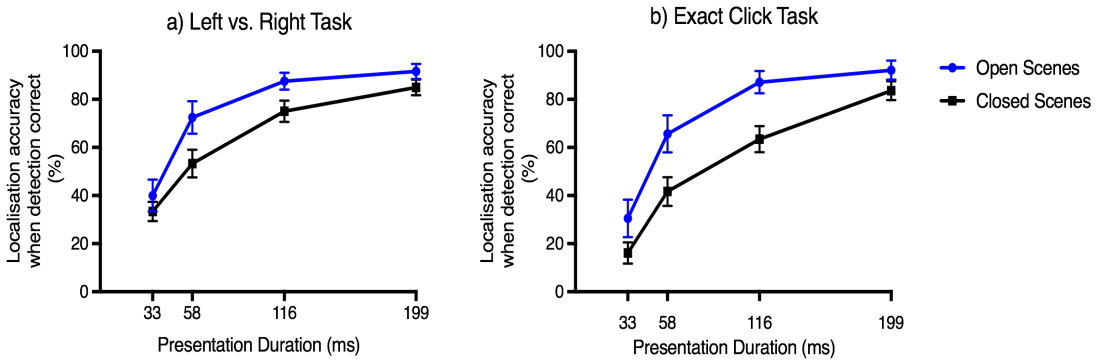


Figure 8. Localisation accuracy as a function of exposure duration partitioned into open vs. closed scene categories in a post-hoc analysis. Note that only trials on which the target was correctly detected are included. The error bars represent 95% confidence intervals.

Experiment 2

Experiment 2 was designed to systematically investigate the influence of clutter in natural scenes on target detection and localisation. As outlined in Experiment 1, we manipulated visual clutter using the computational definition of *enclosure* for natural scenes with the categories: open (coast, open country and highway) and closed (forest, mountain and city; Oliva & Torralba, 2001). The post-hoc analysis of clutter for Experiment 1 was limited due to the small and unbalanced set of open and closed natural scenes across durations, as it was not designed for this purpose. Thus, in Experiment 2, we increased the number of scenes in each category to be equal across open/closed scene types in each duration and examined detection and localisation performance as a function of duration and scene type.

Method

All measures and conditions are reported.

Participants

Thirty participants (24 females, age range 18-58 years, mean [*M*] = 28 years, standard deviation [*SD*] = 8.5) were recruited from Macquarie University. Five had participated in Experiment 1 (15 months prior). All gave informed consent, reported normal or corrected-to-normal vision, and were reimbursed for their time. The study was approved by the Macquarie University Human Research Ethics Committee (Medical Sciences).

Stimuli

There were 240 natural scene images in total. Half of the images of scenes were repeated from Experiment 1 and half were new images selected from the Web using Google Images to reach a balanced number across the open/closed categories. The open (coast, open country and highway) images and closed (city, forest and mountain)

contained 40 exemplars in each basic category (see Fig. 9). The images were converted to grayscale, downsized and subtended 23° x 15°.

Figure 9. Example scenes for the open and closed categories.

Procedure

The stimuli and setup were identical to those of Experiment 1, with the following modifications.

Target detection and localisation task: We used a Scene (open, closed) x Duration (33 58, 116, 199ms) within-subjects design. To reduce the effects of possible keyboard assignment we changed the detection response keys to a one hand response; ‘left arrow’ for Scene 1 and ‘right arrow’ for Scene 2. We constrained possible target locations to be one of 12 locations, which were equally prevalent. Following the detection response, participants were presented with a screen with a grid of 12 squares and asked to indicate the location of the target using the mouse to select one of the 12 possible locations. Chance could therefore be calculated precisely as 8.3% (1/12). For each interval within a

trial, the scenes were randomly selected from within the same open or closed category. Duration order was counterbalanced across participants. After 12 practice trials with feedback (3 at each duration), there were 240 experimental trials (60 at each duration; 30 trials per condition (See Fig. 10).

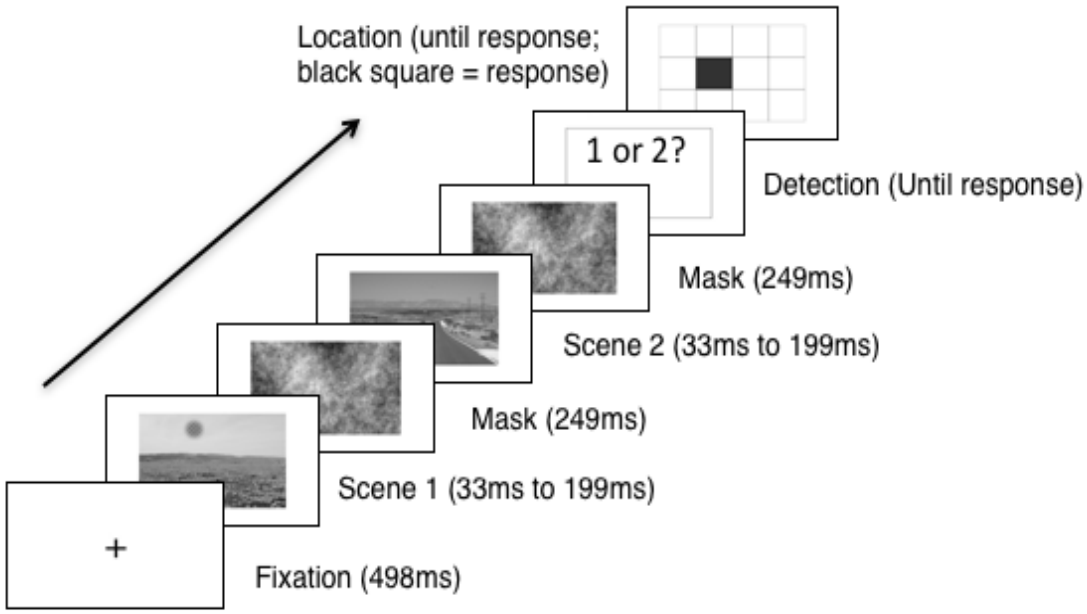


Figure 10. Example of a trial in Experiment 2 for the target detection and localisation tasks. A Gabor target is in Scene 1 of an open scene.

Results and Discussion

Detection performance

First, we evaluated whether there were any differences in detection performance between the open and the closed scenes. Figure 11 shows that performance was higher for the open than closed scenes and this difference increases with longer duration. We calculated d′ as a measure of detection performance and conducted a two-way repeated-measures ANOVA with the factors of Scene (Open, Closed) x Duration (33, 58, 116,

199) on the mean d′ values. There was a main effect of Scene, $F(1, 29) = 202.24$, $p = < .0001$, $\eta^2p = .88$, a main effect for Duration, $F(3, 87) = 146.42$, $p < .0001$, $\eta^2p = .84$, and a significant Scene x Duration interaction, $F(3, 87) = 10.86$, $p < .0001$, $\eta^2p = .27$. A Bonferroni correction for multiple comparisons ($\alpha = 0.05/4 = .0125$) was applied to post-hoc analyses. The interaction was due to d prime being significantly higher in the open compared with the closed scenes at all durations ($p < .0001$) except for 33ms ($p = .078$) (See Fig. 11).

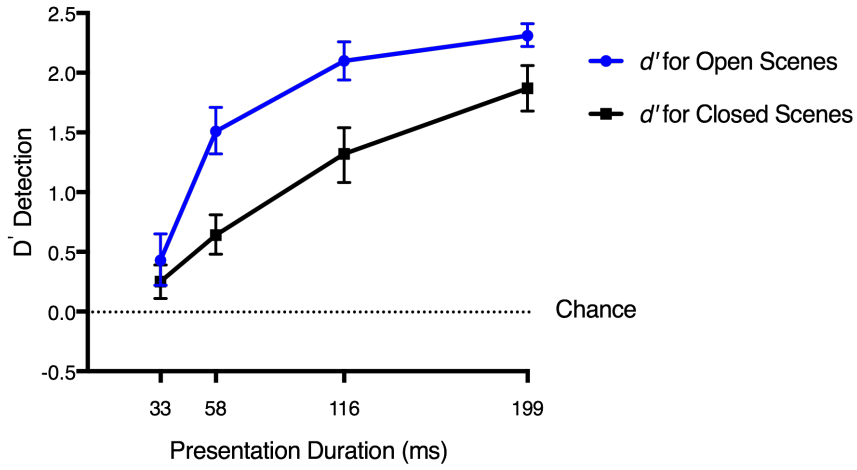


Figure 11. Detection performance for open versus closed scenes. The dashed line represents chance. The error bars represent 95% confidence intervals.

We next examined whether localisation performance was influenced by scene type. As for Experiment 1, trials were partitioned into those in which detection and localisation were correct or incorrect to examine the response distribution as a function of duration and scene type. Figure 12 shows the total proportion of trials across all four durations when detection and localisation were both correct (dark blue bar), when detection was correct and localisation was incorrect (light blue bar), when localisation

was correct and detection incorrect (dark grey bar) and when detection and localisation were both incorrect (light grey bar). Each experimental trial is represented once in the graph. The proportion of trials on which both detection and localisation were correct increases as a function of duration. For the open scenes, numerically there are more trials in which the target is both detected and localised across all durations, however we reserve the quantitative comparison for the following analysis performed on all localisation correct trials.

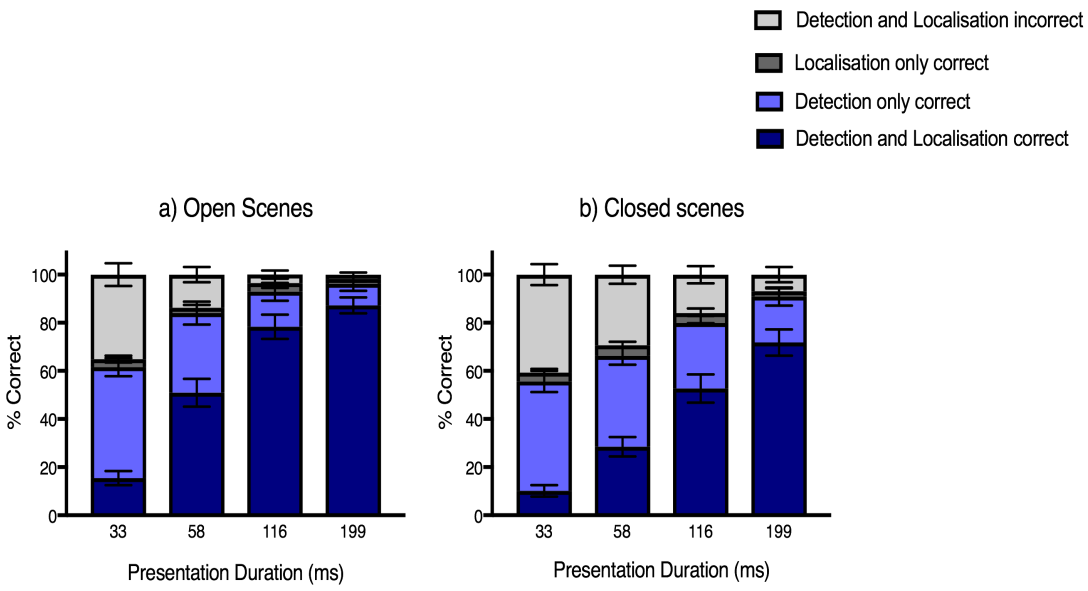


Figure 12. Percentage of trials across all four durations grouped as a function of response profile: trials on which detection and localisation were both correct (dark blue bar), trials on which detection was correct and localisation was incorrect (light blue bar), trials on which localisation was correct and detection was incorrect (dark grey bars) and trials on which detection and localisation were incorrect (light grey bar) for (a) Open scenes and (b) Closed scenes. The error bars represent 95% confidence intervals.

Localisation performance

To investigate the degree to which a briefly presented target can be spatially located when embedded in open versus closed scenes, we analysed accuracy for localisation across all trials, regardless of whether detection was correct or incorrect. Figure 13 shows that localisation accuracy appears higher for the open compared with the closed scenes and this improves with duration. We conducted a two-way repeated measures ANOVA with the factors of Scene (Open, Closed) x Duration (33, 58, 116,

199) on the mean percentage localisation accuracy values. There were significant main effects of Scene, $F(1, 29) = 328.48$, $p = < .0001$, $\eta^2_p = .92$, and Duration, $F(3, 87) = 374.88$, $p < .0001$, $\eta^2_p = .93$, and a significant Scene x Duration interaction, $F(3, 87) = 16.43$, $p < .0001$, $\eta^2_p = .36$. As in the Experiment 1 exploratory post-hoc analysis of clutter, the interaction shows that scene category (open vs. closed) has a different effect on localisation accuracy depending on duration. Although participants were able to localise a salient target with high levels of accuracy for both the open and closed scenes from around 116ms, they were more accurate in both detection and localisation for open scenes compared with closed scenes. Using a Bonferroni correction ($\alpha = 0.05/4 = .0125$), paired sample t- tests showed that localisation accuracy was significantly higher in the open compared with the closed scenes at all four durations [33ms, $t(29) = 3.23$, $p = .003$; 58ms, $t(29) = 7.42$, $p < .0001$; 116ms, $t(29) = 13.48$, $p < .0001$; 199ms, $t(29) = 8.97$, $p < .0001$]. To replicate the findings of Experiment 1, we also conducted simple sample *t*-tests (Bonferroni corrected) on localisation accuracy. Compared to a chance performance of 8.3%, observers were above chance at all durations for both open scenes and closed scenes [Open scenes: 33ms, $t(29) = 8.88$, $p < .0001$; 58ms, $t(29) = 16.06$, $p < .0001$; 116ms, $t(29) = 34.98$, $p < .0001$; 199ms, $t(29) = 54.1$, $p < .0001$; Closed scenes: 33ms, $t(29) = 3.56$, $p = .001$; 58ms, $t(29) = 11.48$, $p < .0001$; 116ms, $t(29) = 18.38$, $p < .0001$; 199ms, $t(29) = 30.23$, $p < .0001$]. Even at the briefest duration, localisation performance was above chance for both the uncluttered (open) and cluttered (closed) scenes, with a steady rise in performance with increasing presentation duration. These results demonstrate that detection and localisation performance is affected by visual clutter.

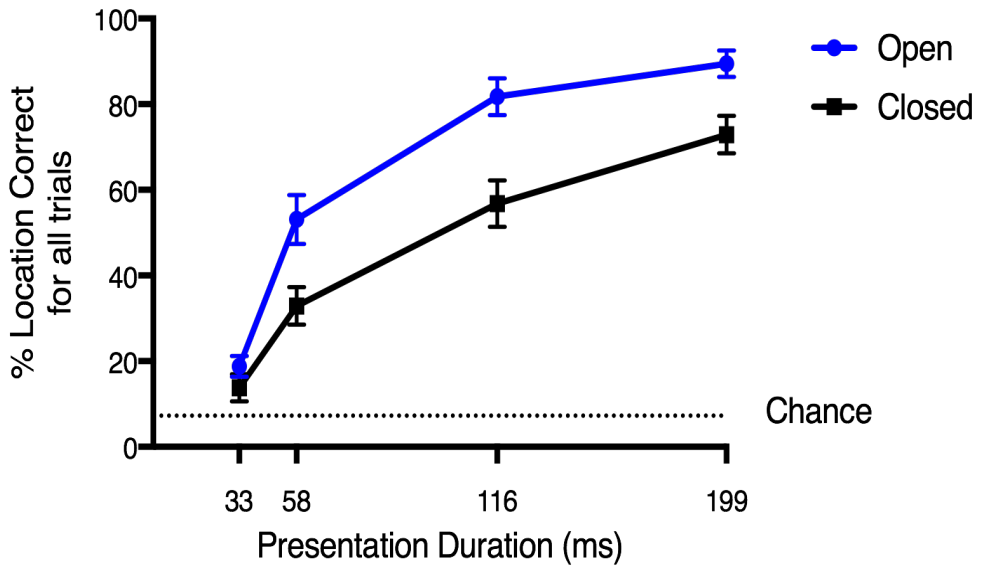


Figure 13. Localisation performance for open and closed scene categories. All trials are included in the analysis, regardless of whether target detection was correct. The dashed line represents chance (8.3%). The error bars represent 95% confidence intervals.

General Discussion

Observers can report the basic category of a scene (e.g., natural or man-made) and detect targets in both natural scenes (Green & Oliva, 2009; Potter & Faulconer, 1975; Thorpe et al., 1996; VanRullen & Thorpe, 2001) and medical images (Evans et al., 2013; Evans et al., 2016; Kundel & Nodine, 1975), at very brief display durations. Although location information is crucial for guiding our interactions with the environment, some findings suggest that detecting a target and localising it might not always go together (Evans et al, 2013; Evans et al., 2016; Howe & Webb, 2014). Here, we used natural scenes and showed observers can both detect and localise a Gabor target at durations as brief as 33ms.

In Experiment 1, even for the most difficult exact click localisation task, which requires precise location information, observers had sufficient information from a 33ms presentation to be able to perform greater than chance. This suggests that fine-grained localisation information can be extracted in as little as 33ms. The results of Experiment 2 showed that this is mediated by visual clutter, with increasing levels of visual clutter impairing both detection and localisation performance. Together, our results suggest that at least some localisation information is available at the brief exposure durations that support rapid target detection.

Although the results show that a target in a natural scene can be localised with brief displays, this was not the case on all trials; performance was far from ceiling. Additionally, there are a proportion of trials where detection was correct and localisation was incorrect, especially at the briefest durations (33 and 58ms) and in the more precise exact click localisation task. The challenge in interpreting these particular results is that there are several possible interpretations. These trials may truly reflect a dissociation between information for detection and information for localisation, as has been suggested in the medical literature, where target detection but not localisation was above chance at 500ms (Evans et al. 2013). However, apparent detection without localisation could also occur for a number of other reasons. First, on these trials participants may have guessed the correct response (was the target in Scene 1 or Scene 2?), without actually having seen the target on that trial. These 'lucky guesses' would presumably be accompanied by a guess on the subsequent localisation, which has only 1/12 chance of being correct. Second, observers may have had some localisation information on these trials, but this was too coarse to be detected by the localisation tasks. This is suggested by the greater occurrence of apparent 'detection without localisation' trials for the more precise exact click localisation task compared to the left/right localisation judgment. Finally, there may

have been localisation information present on these trials but our experiment was not sensitive enough to detect it. We found that the proportion of 'detection without localisation' trials decreased with increasing exposure duration, but this could be consistent with either the correct guessing account or a dissociation of detection versus localisation information. To make the claim of a dissociation between detection and localisation, one would need to correct for guessing, and then conduct a Bayesian analysis to test whether there is evidence for a *lack* of location information.

Our main findings demonstrate that a briefly viewed scene could be processed sufficiently for a target to be localised. In our experiments, for some scenes, the Gabor would be salient (e.g., the open scenes, Figure 13), guiding attention effectively. Performance was far from ceiling, however, suggesting that we have not reached the realm of 'pop-out' in these displays (Borji & Itti, 2013; Rosenholtz et al., 2007). For other scenes (e.g., the closed scenes), there was less salient information about the target, but some information about localisation was still available. Using this paradigm allowed us to precisely control for low-level visual factors (e.g. target size, contrast, spatial frequency) as well as semantic congruency. Although we found localisation of a target in natural scenes with rapid viewing, it is possible that either our displays were not brief enough to dissociate detection and localisation due to the speed of visual processing of natural scenes, or that the effects seen in complex medical images with subtle masses (e.g., Evans et al., 2013; Evans et al., 2016) do not generalise to natural scenes. In addition, the alternative interpretations we offer above for why target detection may occur without localisation on some trials in our own results also apply equally to the studies with medical images.

The results of Experiment 2 are consistent with what we know about visual search in free viewing, with increasing clutter or set size decreasing performance (Adamo, Cain

& Mitroff, 2015; Asher, Tolhurst, Troscianko, & Gilchrist, 2013; Rosenholtz, Li, Mansfield & Gin, 2005; Rosenholtz et al., 2007). Reduced performance with increased clutter may be due to crowding and/or masking effects by the increased number of items in the set size of scene contents. These effects have been documented throughout the visual search literature (e.g., Asher et al., 2013; Whitney & Levi, 2011; Wolfe, 1994). The reduced accuracy we observe for scenes with increased clutter is consistent with the idea that the task would be more difficult for some medical images, such as highly-cluttered or dense breast tissue.

The relationship between clutter and target detection and localisation has potential real-world implications. Analogous to clutter interfering with performance in natural scenes, in which we are all experts, it seems likely that similar effects occur for radiologists interpreting medical images, for example during mammography screening. In the medical perception literature, there have been a number of studies that have investigated factors such as lesion subtlety, which may be dependent on the surrounding anatomical structures (e.g., Krupinski, 2005). Female breast tissue is extremely variable with regards to mammographic breast density (Li, et al., 2013). This dense tissue can increase the visual complexity for a radiologist, potentially masking and/or distracting from pathology. As an extension to the present work, we are currently conducting similar experiments in our laboratory on medical experts. To explore the effect of visual clutter in an applied medical context we are using a related task with radiologists to investigate abnormality detection in mammograms of different density. This will provide an important extension to the current finding that localisation of a target in a complex visual scene is possible at very rapid exposure durations and is influenced by clutter.

Our rapid perception of the world around us is critical for successful interactions. Here, we explored the degree to which information available in very brief presentations

of natural scenes could support not just detection of a target but also knowledge of where that target is. Access to location information is crucial for guiding actions or further analysis (e.g., eye movements); we find a tight link between information supporting detection and localisation.

References

Adamo, S.H., Cain, M.S., & Mitroff, S.R. (2015). Targets need their own personal space: Effects of clutter on multiple-target search accuracy. *Perception*, *44*(10), 1203–1214. http://doi.org/10.1177/0301006615594921.

Asher, M.F., Tolhurst, D.J., Troscianko, T., & Gilchrist, I.D. (2013). Regional effects of clutter on human target detection performance. *Journal of Vision*, *13*(5)*:25*, 1-15.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, *35.1,* 185–207.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. doi:10.1163/156856897X00357

Bravo, M.J., & Farid, H. (2007). The depth of distractor processing in search with clutter. *Perception, 36*(6), 643 – 652.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559-564.

Evans, K.K., Georgian-Smith, D., Tambouret, R., Birdwell, R. L., & Wolfe, J.M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin Review, 20*(6), *1170-1175.*

Evans, K.K., Haygood, T.M., Cooper, J., Culpan, A.-M., & Wolfe, J.M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National*

Academy of Sciences, *113*(37), 10292–10297. http://doi.org/10.1073/pnas.1606187113.

Greene, M.R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137-176.

Howe, P.D.L. & Webb, M.E. (2014). Detecting unidentified changes. *PloS ONE, 9*(1), e84490.

Li, J., Humphreys, K., Eriksson, L., Edgnen, G., Czene, K., & Hall, P. (2013). Mammographic density reduction is a prognostic marker of response to adjuvant tamoxifen therapy in postmenopausal patients with breast cancer. *Journal of Clinical Oncology, 31*(18), 2249-2256.

Joubert, O.R., Rousselet, G.A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorisation and object interference. *Vision Research, 47*(26), 3286 – 3297.

Krupinski, E.A. (2005). Visual search of mammographic images: Influence of lesion subtlety. *Academic radiology*, *12*(8), 965-969.

Kundel, H.L., & Nodine, C.F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*(3), 527-532.

Kundel, H., Nodine, C., Conant, E. F., & Weinstein, S.P. (2007b). Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology, 242*(2)*,* 396-402.

Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., Stalder, S., & Maitz, G. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Academic Radiology, 12*(7)*,* 830-840.

Mitroff, S.R., & Simons, D.J. (2002). Changes are not localized before they are explicitly detected. *Visual Cognition*, *9*(8), 937–968. http://doi.org/10.1080/13506280143000476.

Oliva, A. (2005). Gist of the scene. *Neurobiology of Attention*, *696*(64), 251–258.

Oliva, A., Mack, M.L., Shrestha, M., & Peeper, A. (2004, August). Identifying the perceptual dimensions of visual complexity of scenes. In *Proc. of the 26th Annual Meeting of the Cognitive Science Society* (Vol 26, No. 26, pp. 101-106).

Oliva, A. & Torralba, A. (2001). Publication utilizing eight semantically organized categories: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. From http://cvcl.mit.edu/database.htm

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Potter, M.C. (1976). Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem. 2*(5), 509–522.

Potter, M.C., & Faulconer, B.A. (1975). Time to understand pictures and words. *Nature,* 253, 437–438.

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005, April). Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 761-770). ACM.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17–17. http://doi.org/10.1167/7.2.17.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582*),* 520 – 522.

VanRullen, R., & Thorpe, S.J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artefactual objects. *Perception, 30*(6), 655-688.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, *15*(4), 160-168.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.

Chapter 3: Finding cancer in mammograms: if you know it's there, do you know where?

Chapter 3: Finding cancer in mammograms: if you know it's there, do you know where?

Chapter 3 is comprised of two experiments. The first experiment was a pilot study and the second experiment extends and builds upon these findings. As such, for this chapter, the introduction and discussion have been modified to encompass both experiments and to avoid repetition. The manuscript that describes Experiment 2 (modified slightly for this Chapter) was co-authored by Carrigan, A.J., Wardle, S.G., & Rich, A.N. (under review). Finding cancer in mammograms: if you know it's there do you know where it is? *Cognitive Research: Principles and Implications*. The version submitted for publication can be found in Appendix A.

As soon as we open our eyes, our visual system processes an enormous amount of information in a short space of time. Early findings showed that an exposure of 100ms is sufficient to extract the basic meaning of natural scenes (e.g., indoor versus outdoor; Potter, 1976). Using backward masking to precisely control for exposure times, others have shown that the distinction between natural scene categories at the superordinate level (e.g., man-made versus natural) and basic level (e.g., coast versus city) can occur with presentation durations as short as 20ms (Greene & Oliva, 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007). Furthermore, when observers are pre-specified a category (e.g., animal or truck), objects can be detected at brief durations (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). This fast-visual processing has also been reported among those who are experienced in domain-specific tasks such as medical imaging (Evans, Georgian-Smith, Tambouret, Birdwell & Wolfe, 2013; Evans, Haygood, Cooper, Culpan & Wolfe, 2016; Kundel & Nodine, 1975; Nodine, et al., 1999). Kundel & Nodine (1975) showed that when presented a chest radiograph for 200ms, radiologists could detect an abnormality with 70% accuracy. Kundel and colleagues (2008) have since shown that within 1 second of viewing a mammogram, experts fixate on 67% of breast cancers (Kundel, Nodine, Krupinski, Mello-Thoms, 2008). Furthermore, when shown briefly presented mammographic displays (250ms), radiologists can discriminate normal from abnormal at levels better than guessing (Evans et al., 2013; Evans et al., 2016). The evidence that observers can extract information with fast presentations from natural scenes (e.g., Potter, 1976; Thorpe et al., 1996), and medical images (e.g., Kundel & Nodine, 1975; Evans et al., 2013), suggests that the processing involved in early visual search is similar whether the display is a natural scene or a medical image, at least for experts.

Radiologists develop expertise in 'visual search' in such images over a period of years. It has been suggested that specialised training and ongoing experience leads to perceptual and cognitive 'fine-tuning' in the task of image interpretation (Nodine & Mello-Thoms, 2010). Maintaining such expertise requires interpreting high volumes of cases. For example, mammographic screening radiologists interpret more than 2000 cases per year (Rawashdeh, et al., 2013). There is evidence that this extensive experience modulates the perceptual/cognitive system of experts: experienced radiologists outperform novices and trainee radiologists on tasks such as detecting an abnormality in brief images (Evans et al., 2013; Nodine et al., 1999), and in different patterns of eye movements between experts and novices. For example, Kundel and La Follette (1972) compared the visual scan patterns of expert breast radiologists with trainees interpreting mammograms and found that the experts fixated on lesions faster and concluded search earlier than the novices. Others have shown that experts fixate true abnormalities within 1-2 sec of image onset and most of their subsequent scanning is to confirm that there are no other lesions (Mello-Thoms, et al., 2005). This follow-up takes about 5-10 seconds after initial fixation, after which a diagnostic decision is reached. There is an enormous amount of information that is processed in the first second of viewing a scene or image, so it is important that we understanding the cognitive underpinnings of early visual search.

Kundel and Nodine (1975) developed a model that describes two distinct processes leading to a diagnostic decision. The first glance supports a global, or holistic overview of the image, which indicates on a basic level whether the image deviates from a cognitive representation of a normal anatomical schema. The information extracted at this first stage is then proposed to constrain and guide search to the region of the image

containing the abnormality (the second stage). For this to occur, the global signal must be informative about the location of the abnormality.

Recently an alternative perspective has been offered by Evans and colleagues (2013; 2016). They suggest an initial abnormal signal could act to alert a radiologist that *something* is abnormal but without containing location information. Rather than guiding search to a location, this global signal then *changes the search strategy* to a more complete search for the abnormality. The initial signal could be supported by the rapid extraction of the summary statistics of the image, such as average orientation and size. In the basic vision literature, two stage models (e.g., Wolfe, Võ, Evans & Greene, 2011) describe an initial, non-selective pathway which, although limited in capacity, extracts summary statistics in parallel from the display. In the model, global processing occurs along this pathway. A second, selective pathway recognises one or a few objects at a time and requires selective attention. Together these pathways combine to support perception. Evans et al. (2013) and Evans et al. (2016) suggest that information via the non-selective pathway could alert a radiologist that something is abnormal, but the fine-grained detail, such as its location, only becomes available at the later selective stage.

Evans et al. (2013) compared the performance of radiologists and novices on the detection and localisation of abnormalities in mammograms. The stimuli were bilateral (left and right breast) mammograms where one side could contain subtle masses and architectural distortions that varied in size (10 to 48mm). Such pathologies are highly variable, and are difficult to detect and locate even by expert radiologists under free viewing conditions. As a result, these have the highest reported rate of false negatives (Knutzen & Gisvold, 1993). Despite these difficult images, Evans et al. (2013) found that radiologists (but not novices) could detect an abnormality above chance (Mean d' was ~ 0.7 for 250ms duration and up to ~1 for 2000ms duration, where d' of 0 is chance). For

the combined detection and localisation task, images were displayed for 500ms. Following detection, the radiologists viewed a blank outline of the mammogram and were asked to localise by marking the abnormality with a mouse-click. Chance was determined by calculating the average percentage (across images) of overall tissue area lying within a predetermined region of abnormality. Although abnormalities could be detected by radiologists above chance at 500ms, localisation performance was at chance. Evans et al. (2013) interpreted these results as evidence that the information extracted to support detection at brief durations does not contain location information, but is rather based on an overall 'gist' or holistic signal.

In a subsequent paper, Evans et al. (2016) did another series of experiments using mammograms, replicating and extending their initial findings. In their second experiment, they presented radiologists a set of 120 single-sided (one breast) mammograms for 500ms and asked them to detect and then localise an abnormality. The unilateral mammograms either contained an abnormality (target-present), had no abnormality (target-absent), or was the contralateral breast from the target-present mammogram (no abnormality). In this experiment, mean d' for detection was 1.16 for the target-present/target-absent images, significantly above chance (0), whereas localisation accuracy was not significantly greater than that expected by chance (6%). They concluded that the radiologists could not localise a lesion despite detecting it. Further, they suggested that experienced radiologists could even make such judgments based on images from the contralateral (thus far normal) breast (remaining 40 images). Mean d' was 0.59 for detection of abnormality in the contralateral breast from a woman with signs of cancer in the other breast. This result is striking because the mammogram on which the judgement was based had no mass. These results provide intriguing hints that the information required for detection and that for localisation could be dissociable, and indeed that even images without an actual mass

may carry some signal that is informative to experts about the overall potential for cancer in the patient.

Evans et al. (2013) and Evans et al. (2016) interpret their results as reflecting a global signal of abnormality that lacks information about location of a specific mass. Indeed, the remarkable findings that a diagnosis could be made from the contralateral apparently-normal breast when the opposite side was abnormal might be explained by this interpretation. There are, however, some alternative interpretations that need to be carefully considered and ruled out. Frequentist statistics, used in these studies, cannot distinguish between a true null (no effect exists) and a lack of sensitivity (an effect exists but is not detected). To interpret a null effect as evidence for there being no effect (in this case no localisation), we would need to use alternate statistics, such as a Bayes Factor (Dienes, 2011). Second, the summary statistics (e.g., average *d* prime) are inadequate to answer the key questions. For *d* prime values quite close to chance, artefacts or slight imprecisions in localisation for just a few images could be sufficient to drive performance to an apparently greater than chance level. For example, if participants are actually 'detecting' a distracting signal in the breast for a target present trial, the detection response would be correct, but localisation would be incorrect (as participants would click on the distractor). Similarly, if participants click just outside the lesion, this would be categorised as incorrect, which would lead to the erroneous inference that there was no localisation information. Finally, in a detection experiment, there will always be some 'lucky guesses' that are correct. We need to consider the impact of these on the apparent dissociation between detection and localisation. These two studies by Evans and colleagues (2013) and (2016) raise important questions, but the challenge to the Kundel and Nodine (1975) model of radiologists' diagnostic decision-making rests heavily on the lack of information about the location of an abnormality. We need to go beyond the

summary statistics and explore image level variability, precision of localisation responses and the potential influence of guesses to ascertain whether there is truly detection without localisation.

The aims of the present study were to extend previous work by Evans et al. (2013) and explore the question of localisation in detail. The claim that radiologists can detect an abnormality without knowing where it is has strong theoretical implications. Instead of the intuitive notion that the information in the first glance guides attention and the eyes towards the location of the potential abnormality, it implies a quite different process. Here, our first aim was to see whether expert readers of mammograms viewing brief displays can extract location information when there are localisable signals from the lesion (i.e., a salient mass).

Experiment 1: Pilot study to validate the paradigm

This pilot experiment was designed to test whether experienced radiologists could extract sufficient information about obvious masses in a single mammogram to support detection and localisation at brief presentation durations. We presented a single side (unilateral breast) to reduce the size of the visual display and amount of visual processing required (relative to Evans et al (2016, Experiment 2)). We also were meticulous in selecting our stimuli. Using real-world stimuli rather than typical laboratory visual search displays allows for high ecological validity, but presents considerable challenges. The available images are often far from perfect for vision studies as it is difficult to control for factors such as co-existing variables (e.g., breast calcifications, target number and size, and breast tissue type). For our experiment, we selected images with single masses and minimised the features in the image which may act as visual distractors for the radiologists.

We investigate detection and localisation performance for a single mass present in a unilateral mammogram at unmasked exposure durations of 250ms, 500ms and 750ms. The majority of the stimuli were categorised as ‘masses easy to detect’ by an independent breast radiologist (P.S) on free viewing. The participants performed a detection and an ‘exact click’ localisation task similar to Evans and colleagues (2013), except that the target-present stimuli exclusively contained a single mass rather than ‘subtle masses and architectural distortions’. We predict that these easy cases (obvious mass) will result in both detection and localisation even at brief durations, demonstrating the task is appropriate for measuring location information when it is able to be extracted.

Method

All measures and conditions are reported.

Participants

We defined experts as having at least four years of experience and in their current practice reading at least 2000 mammographic cases per year (Rawashdeh et al., 2013). When conducting research on experts, the sample size is often constrained by the availability of the participants. Here, we recruited 30 radiologists, and analysed the data from 18 of these who met our criteria for experts. These were all board-certified radiologists (nine females; age range 41 – 73 years, mean [*M*] = 54, standard deviation [*SD*] = 8.6). Half of the participants were NSW BreastScreen radiologists who read > 5000 cases per year and half were from general radiology practices. All gave informed consent and reported normal or corrected-to-normal vision. The study was approved by the Macquarie University Human Research Ethics Committee (Medical Sciences).

Design, Stimuli and Apparatus

The experiment was a single factor repeated-measures design. The stimuli were 96 full-field, de-identified, medio-lateral oblique digital breast mammograms obtained from the Dokuz Eylul Mammography Set (DEMS: Bulu, Alpkocak & Balci, 2013). Half of the images contained a single suspicious mass and half were normal. The average size of the mass was 26.70 mm (*SD* = 13.23 mm) and the range was from 8 – 54 mm. The abnormal images contained a mass previously diagnosed and coded according to the Breast Imaging and Reporting Data System (BIRADS: American College of Radiology: Breast Imaging Reporting and Data System Atlas. Reston, Va: © American College of Radiology, 2003). BIRADS is a standardised breast assessment tool developed for mammography that ranges in number from zero to six. In clinical practice a radiologist assigns a BIRADS score to each image, which determines the next step in the diagnostic protocol. The 52 normal images had assigned BIRADS codes of 1 (no significant abnormality). The abnormal breast images consisted of a mixture of BIRADS codes 4 (suspicious abnormality and biopsy recommended), 5 (highly suggestive of malignancy) and 6 (known pathological proven malignancy). The starting resolution of the single mammograms were 4096 x 3328 or 3328 x 2560 pixels and were downsized to 50° x 30° or 50° x 28° visual angle.

The participants sat at a viewing distance of 53cm in a dimly lit, quiet room at a conference setting. The experiment was written and presented via MATLAB 2013B using the Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997) and controlled using a Macintosh MacBook Pro. The stimuli were displayed externally on a Dell Full HD LCD 22-inch screen with resolution = 1920 x 1200, refresh rate = 120 Hz.

Procedure

We presented the stimuli at one of three presentation durations (250ms, 500ms or 750ms) in separate blocks, counterbalanced across participants. After 4 practice trials at each duration, with images not used in the main experiment, radiologists then completed 3 blocks, each with 32 unique mammograms. Thus, there were 32 trials for each duration. For each participant, the particular image presented in each duration was randomly selected without replacement. Figure 1 shows the trial sequence. Each trial began with a fixation point for 500ms, followed by a centrally-presented left medio-lateral oblique (MLO) breast image. After the mammogram, we presented a black screen asking the radiologists to categorise the mammogram using a key press as either ‘normal’ or ‘contains a suspicious mass’, followed by a black screen with a white mask of the breast (each unique mammogram was paired with its corresponding mask). The radiologists were asked to click with the mouse on the location where they saw a mass. In the case of normal responses, they were asked to click anywhere on the display. They began the next trial with a key press. Participants did not receive feedback and were not told the prevalence of abnormalities (See Fig. 1).

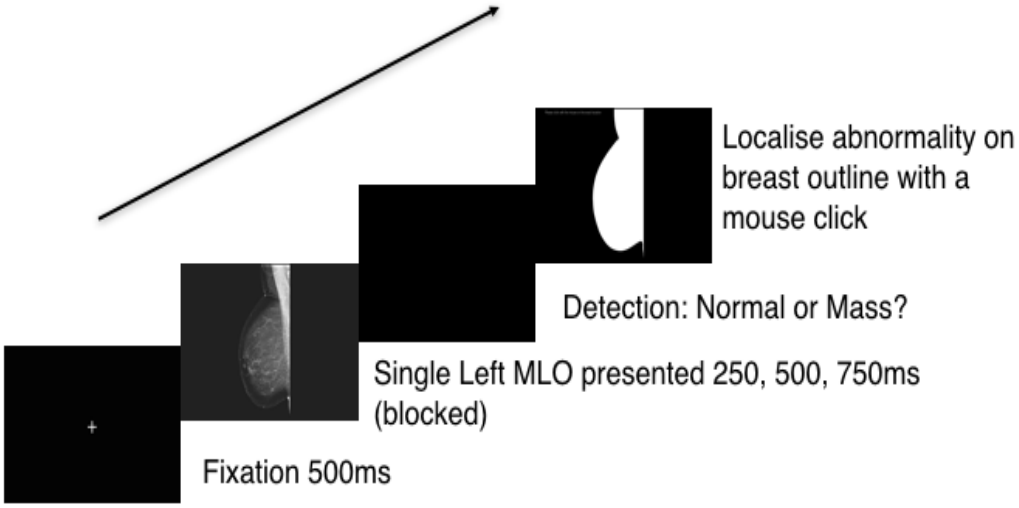


Figure 1: Example trial for Experiment 1. Radiologists were asked first whether the image was normal or contained a mass, and then to use the mouse to indicate the location of the mass if present. MLO = Medio-lateral oblique mammographic projection.

Analysis

Following the recommendations of Cumming (2012), we present Mean differences (M_{diff}) with 95% confidence intervals (CI), as well as a Cohen's d estimate of effect size corrected for small sample size, to assist in accurate interpretation of the effects. This latter measure, d_{unb} represents an adjusted, unbiased Cohen's *d* standardised effect size applied to single sample t-tests where d_{unb} = (1 - 3 / (4*df - 1)) * *d* (Cumming, 2012). A Bonferroni adjustment was applied to all statistical tests (alpha = .0167).

Results and Discussion

Ten images were removed post hoc (five from the cancer set and five from the normal set) prior to the analysis as these were found to contain co-existing pathology that may have confounded the results (e.g., microcalcifications). The responses from the participants for a total of 86 images were analysed.

Detection accuracy: Sensitivity (*d′*) was calculated as a function of abnormality present or absent. Higher *d′* indicates greater sensitivity: the higher the *d'*, the more accurately the radiologists gave the correct answer on both target present and target absent trials (i.e., reported a mass when a mass was present *and* no mass when no mass was present). A *d′* of zero indicates there is no sensitivity and the participant is performing at chance (i.e., no better than guessing). In free-viewing tasks, radiologists have been shown to have *d′* values around 2.5–3.0 (D'Orsi, et al., 2013). Figure 2a shows the average *d′* values across participants for the three durations. Single sample t-tests showed the radiologists' detection was significantly above chance [t-test relative to 0: 250ms, $t(17) = 11.81$, $p < .0001$, $M_{diff} = 1.28$, CI [1.05, 1.50], $d_{unb} = 3.76$; 500ms, $t(17) = 12.32$, $p < .0001$, $M_{diff} = 1.67$, CI [1.39, 1.96], $d_{unb} = 3.92$; 750ms, $t(17) = 10.53$, $p < .0001$, $M_{diff} = 1.44$, CI [1.15, 1.73], $d_{unb} = 3.35$]. Repeated measures ANOVA on *d'* revealed no effect of presentation duration [$F(2,34) = 2.28$, $p = .117$].

Of course, these *d′* values reflect poorer performance than seen in free-viewing – and in clinical practice, additional mammographic projections, previous imaging and patient history would be available to a reporting radiologist. Thus, the task of detecting an abnormality after seeing a one-shot display flash briefly is not the typical way a radiologist conducts image interpretation, but it can give valuable insights into the type of information available in this initial stage of the diagnostic process.

Localisation accuracy: Our key question was whether there is localisation information when detection is correct so we analysed trials where the participants correctly detected an abnormality at each exposure duration (i.e., correct detection target-present trials) to look at the accuracy of the localisation task. We compared the location of the mouse click with the location of the actual mass and coded the response as either

accurate (participant clicked within the boundaries of the mass) or not (any other location). We then compared localisation performance to chance, calculated across the 43 mass-present images as 4%. This is the proportion of breast tissue that contains the mass to the proportion of total tissue, and represents the average number of possible random locations the radiologists could select, taking into account the lesion and image size across all of the target-present images. Figure 2b shows localisation accuracy (percentage correct) on detection correct trials compared with chance (4%). Single sample t-tests showed that the radiologists' localisation accuracy (when detection was correct) for all three durations was significantly above chance [t-test relative to 4: 250ms, $t(17) = 9.93$, $p < .0001$, $M_{diff} = 30.85$, CI [24.3, 37.41], $d_{unb} = 3.16$; 500ms, $t(17) = 12.04$, $p < .0001$, $M_{diff} = 42.46$, CI [35.02, 49.9], $d_{unb} = 3.83$; 750ms, $t(17) = 12.3$, $p < .0001$, $M_{diff} = 38.66$, CI [32.03, 45.29], $d_{unb} = 3.92$].

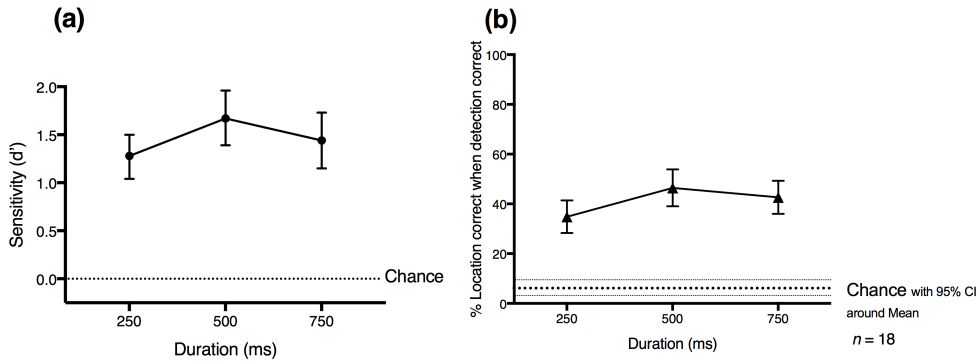


Figure 2: Detection and localisation results for Experiment 1. (a) Average sensitivity (d') on the detection task; (b) Average percentage correct on the localisation task when detection was correct. Chance is 4% (dotted line) with 95% confidence intervals. Error bars represent 95% confidence intervals.

The localisation results show that on at least a third of the trials on which they correctly detected the presence of a mass, participants also had information about the exact location of the mass. To examine localisation performance as a function of detection performance, we binned target present trials according to responses on both detection and localisation. In Figure 3, we present the data in an alternative format to examine the relationship between detection and localisation. Each experimental target-present trial is represented once in the graph. At all durations, there are approximately 20% of trials where the radiologists are correctly detecting a mass without being able to localise it accurately.

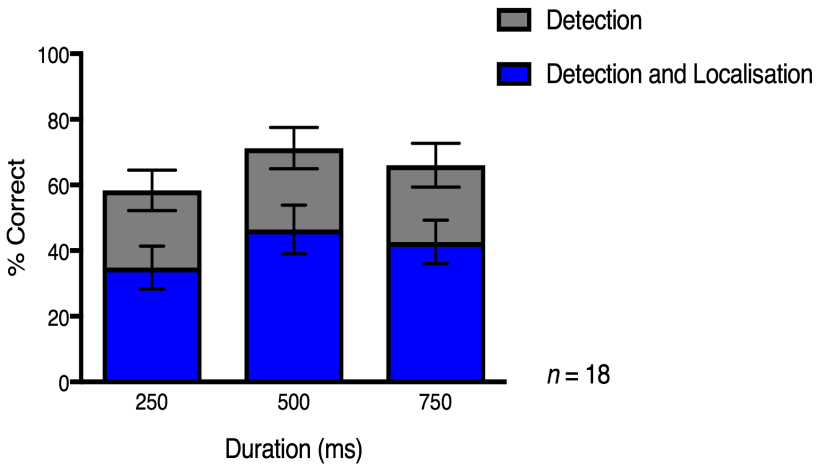


Figure 3: Average percentage correct for detection and localisation correct (blue bar) and detection only correct (grey bar). Error bars represent 95% confidence intervals.

To summarise, in this experiment we tested a group of experienced radiologists on a task designed to measure the information available to them when mammographic images are presented for brief durations. We modified the paradigm used by Evans and colleagues (2013) to test whether localisation information could be sensitively measured in this way, and found a significant portion of trials on which such information was indeed detectable. Radiologists could detect and localise a mass in a single view mammogram which was presented briefly but not masked, as well as some trials where apparently only detection is possible. In these images, 75% of the masses were relatively obvious: the lesion was conspicuous or salient, often due to *low* mammographic breast density (MBD). From this pilot experiment, we designed a follow-up experiment to fully explore whether we can still provide evidence of detection and localisation when the images are more difficult. We manipulated breast density, and tightly controlled presentation durations (using a mask), after which we delve into the 'detection without localisation' data more deeply.

Experiment 2

The aims of Experiment 2 are to extend previous work by Evans and colleagues (2013; 2016) and explore in detail whether detection and localisation are dissociable. Female breast tissue is highly variable in mammographic breast density (MBD: Li, et al., 2013). In the human population, 40% of women aged between 40-74 years have dense breasts (Sprague, et al., 2014). Critically, as MBD increases there is a 4-6-fold increased risk of breast cancer (Boyd, et al., 2010), and studies have shown that higher levels of MBD reduce radiologist sensitivity, thus limiting early detection of breast cancer (Al-Mousa, Ryan, Mello-Thoms & Brennan, 2014). For a radiologist, MBD increases the complexity of the image and could mask and/or distract from existing pathology. Thus, an

additional aim was to explore the effect of breast density *(*which can make masses more difficult to see*)* on the type of information that can be extracted in a brief display. Finally, the distinction between theories of radiologist visual search rests heavily on the dissociation between detection and localisation of masses. Our third aim was therefore to develop methods that can test for evidence of this dissociation. To this end, we looked at the images in detail to explore the degree and source of localisation errors on apparent detection-correct trials, as well as considering the potential influence of 'lucky' guesses to 'detection without localisation' performance.

We investigate detection and localisation performance for a single mass in unilateral mammograms presented centrally for a brief duration and then masked. There is a bias to click directly in front of fixation (centre of the image) when the location is unknown (Buswell, 1935; Tatler, 2007). However, the mass location varied within the breast in our images, which minimises the influence of any such bias (i.e., a random central click is not likely to fall within the mass location on many trials). We presented two sets of mammograms that varied on density (high density and low density) and mass presence. As half of the images contained a mass that would be difficult to detect, we used two durations (unique images in each): 250ms (within the timeframe others have considered to support gist-level information in medical images; Evans et al. (2013)) and 1000ms (presumably well beyond gist level of perception). The participants performed a detection and an 'exact click' localisation task similar to Evans and colleagues (2013), except that the target-present stimuli exclusively contained a single mass that was either easy (50%) or difficult (50%) to see, due to level of breast density, rather than subtle masses and architectural distortions. We predict that mass detection and localisation will be more accurate for mammograms with low density compared with those with high density at both experimental durations. We consider image variability, response

imprecision and we use alternative analyses and a guessing correction to fully test for a dissociation between knowing an abnormality is present versus knowing where it is.

Method

All measures and conditions are reported.

Participants.

Twelve participants with experience in interpreting mammograms were recruited from BreastScreen New South Wales and local radiology practices (6 female, Average age = 54 years, *SD* = 13 years). We defined experts as having at least four years of experience and in their current practice reading at least 2000 mammographic cases per year (Rawashdeh et al., 2013). The BreastScreen doctors (n = 11) read > 3000 mammographic cases per year, but we did also include one breast physician who read > 1000 cases per year, as she had extensive experience (10 years). The average experience reading mammograms of our participants was 22 years (*SD* = 13 years). All gave informed consent and reported normal or corrected-to-normal vision. The study was approved by the Macquarie University Human Research Ethics Committee (Medical Sciences).

Design, Stimuli and Apparatus

We used a Density (low, high) x Duration (250, 1000ms) within-subjects design. As in Experiment 1, the stimuli were 160 full-field, de-identified, medio-lateral oblique digital breast mammograms obtained from the Dokuz Eylul Mammography Set (DEMS: Bulu, et al., 2013), which varied on target presence/absence, and high MBD/low MBD. Half the images (80) were normal and half contained a single mass previously diagnosed and coded according to the Breast Imaging and Reporting Data System (BIRADS: 2003).

The average size of the mass was 26.70 mm (*SD* = 13.23 mm) and the range was from 8 – 54 mm.

Difficulty was manipulated by including two sets of mammograms (dense: high MBD and fatty: low MBD) where half of the mass images (40) and half of the normal images (40) had high MBD. The remaining images had low MBD (See Fig. 4). Density was categorised on a dichotomous scale (low/high) by an experienced radiologist blind to the purpose of the study (M.B.) and one author with experience reading mammographic images (A.C.). These ratings were significantly correlated ($r = 0.9$, $p < .0001$). One of the most challenging aspects of studying radiologists and using medical images rather than using artificial stimuli is that the human body varies widely anatomically. To ensure that the stimuli set used in our experiment was free of potential confounds such as co-existing pathology we included images that only contained a single mass. Image artefacts such as side markers, dust and large calcifications (potential distractors) were found in ten mammograms and were deleted using GraphicConverter (version 9.4).

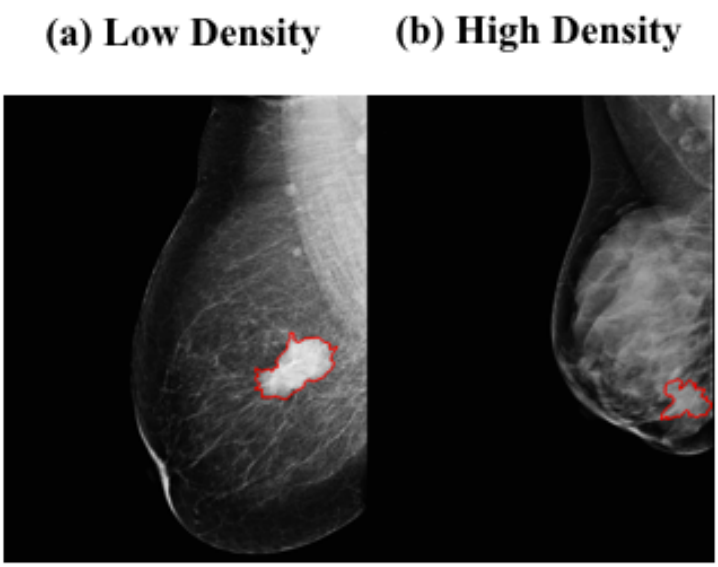


Figure 4: Exemplars of target-present images. The red outline depicts the mass (and did not appear in the actual stimuli). (a) Low density breast that contains predominately fatty tissue, which is radio-translucent or black/grey. The higher contrast mass is easily seen; (b) High density breast that contains normal fibroglandular tissue resulting in a more difficult search. The X-ray beam is attenuated by this tissue and appears radio-opaque or white on a mammogram.

The experiment was presented on a Macintosh MacBook Pro using MATLAB 2011B with the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). The stimuli were centred on a 1920 x 1080 resolution 24-inch, LG W2442PA, liquid-crystal display screen, refresh rate of 120Hz. The participants sat approximately 70cm away from the screen. The original resolution of the single mammograms was 4096 x 3328 or 3328 x 2560 pixels, which were downsized to 19° x 24° (18 out of 160) or 20° x 24° of visual angle. To validate our image categories and presentation durations, pilot data was collected from three radiologists at 250ms and 500ms durations two months prior to their

participation in the experimental session. Previous studies which have used medical images have reported that a time-lapse of around 2 months between each session reduces the likelihood of recall (Berbaum, et al., 2015). On the basis of these pilot data we increased the long duration condition to 1000ms.

Procedure

The experiment was conducted onsite at various metropolitan Sydney BreastScreen and radiology practice locations. We presented the stimuli at two presentation durations (250ms, 1000ms) in separate blocks, counterbalanced in order across participants. For each participant, the particular image presented in each duration was randomly selected without replacement. After four practice trials at 2000ms with feedback and a further six trials at the experimental durations (three at 250ms, three at 1000ms; blocked) with feedback, the radiologists viewed 160 trials without feedback. The radiologists were asked to detect ‘any mass that you would recommend for further investigation’. Each trial began with a fixation point for 500ms, followed by a centrally presented left medio-lateral oblique breast image. This was followed by a backward 1/f noise mask for 250ms after each stimulus presentation and a black screen asking the radiologists to categorise the mammogram using a key press as either ‘normal’ (left arrow key) or ‘mass’ (right arrow key), followed by a black screen with a grey mask of the breast (each unique mammogram was paired with its corresponding mask). The radiologists were asked to ‘please click with the mouse the exact location where you saw a mass’. In the case of normal responses, they were asked to click anywhere on the display. There were 20 trials per condition (duration/target presence/density). Figure 5 shows the trial sequence. Participants began the next trial with a key press.

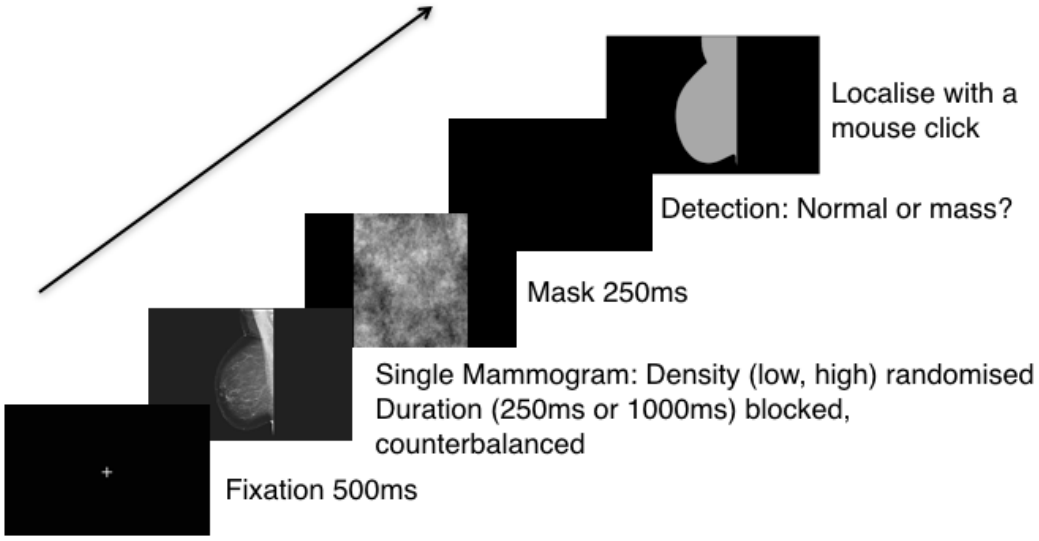


Figure 5: Example trial for twelve radiologists who were asked first whether the image was normal or contained a mass, and then to use the mouse to indicate the location of the mass if present.

Results

The aims of the experiment were to see whether expert readers of mammograms viewing brief displays (1) can extract location information; (2) are affected by breast density in the type of information that can be extracted; (3) show a dissociation between detection and localisation. For all analyses (including Bayes Factor) the Statistical Package for the Social Sciences was used (IBM: SPSS, 2015).

Detection accuracy: Figure 6 shows percentage correct for low density (blue lines) and high density (black lines) for (a) target present and (b) target absent trials. For the target present trials, we can see from Figure 6a that performance for the low density images, in which the masses are salient, appears better than the high density images, where the masses are much more difficult to find even in free viewing. We can also see that accuracy improves with duration. Figure 6b shows accuracy for the target absent

trials. The radiologists appeared less accurate on target absent trials at the longer duration, showing they tended to make false alarms when given slightly more time to inspect the display.

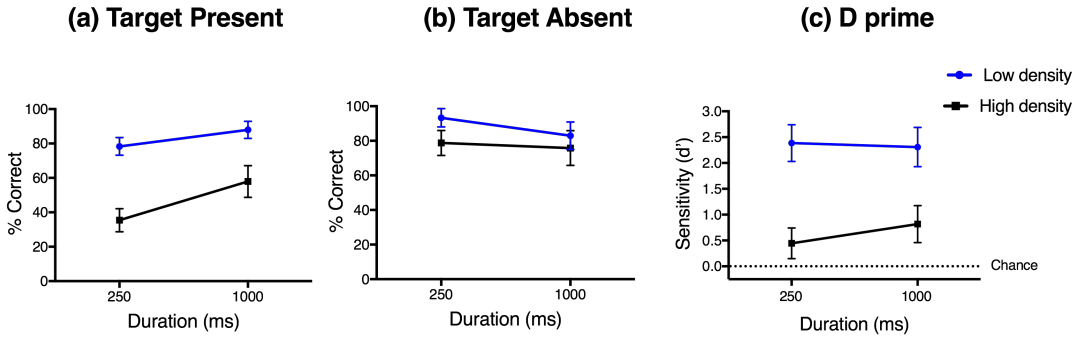


Figure 6: Detection accuracy. (a) Average percentage correct on target present trials; (b) Average percentage correct on target absent trials; (c) Average sensitivity (d′) on the detection task. Error bars represent 95% confidence intervals.

Sensitivity (*d′*) was calculated as a function of abnormality present or absent. Higher *d′* indicates greater sensitivity: the higher the *d′*, the more accurately the radiologists responded to both target present and target absent trials (i.e., reported a mass when a mass was present *and* no mass when no mass was present). A *d′* of zero indicates there is no sensitivity and the participant is performing at chance (i.e., no better than guessing).

Figure 6c presents the sensitivity (*d′*) data. Single sample t-tests (Bonferonni adjusted, alpha = .0125) on average *d′* relative to 0 (chance) for each duration and density showed that radiologists do have information about the presence of the mass at both

durations. Performance at 250ms for the low density condition was greater than chance ($t(11) = 14.97$, $p < .0001$, $M_{diff} = 2.39$, CI [2.03, 2.74], $d_{unb} = 5.69$) as was performance in the more difficult high density images ($t(11) = 3.3$, $p < .007$, $M_{diff} = .44$, CI [.15, .74], $d_{unb} = 1.3$). As one might expect, this was also the case at the longer duration of 1000ms, both for low density images ($t(11) = 13.38$, $p < .0001$, $M_{diff} = 2.31$, CI [1.93, 2.69], $d_{unb} = 5.09$) and high density images ($t(11) = 5.04$, $p < .0001$, $M_{diff} = .82$, CI [.46, 1.17], $d_{unb} = 1.92$). Although high density d' values reflect poorer performance than seen in free-viewing, where radiologists have d' values around 2.5–3.0 (D'Orsi, et al., 2013), performance already approaches these levels for the low density images, even at 250ms (see Fig. 6c). These results suggest that when the mass is relatively easy to see (low density), diagnostic sensitivity in the first quarter of a second is already close to that of free-viewing.

As one would expect, we can see from Figure 6c that performance for the low density images is better than the high density images. This obvious pattern was confirmed by a repeated measures ANOVA with the factors of Density (low, high) x Duration (250, 1000) on the mean d' values. This showed a main effect of Density ($F(1, 11) = 133.51$, $p < .0001$, $\eta^2_p = .92$), no effect of Duration, ($F(1, 11) = .98$, $p = .344$) and no Density x Duration interaction ($F(1, 11) = 2.09$, $p = .18$).

Localisation accuracy: Our key questions were first, whether there is localisation information when detection is correct, and second, how breast density influences localisation. Using the same method as Evans et al. (2013) and Evans et al. (2016, Experiment 2), we compared the location of the mouse click with the location of the actual mass and coded the response as either accurate (participant clicked on or within the boundaries of the mass) or not (any other location). We analysed trials where the participants were correct on detecting an abnormality at each exposure duration (i.e., correct detection target-present trials). We compared localisation performance to chance,

calculated across the 80 target-present images as 4.4% (CI [3.02%, 5.75%]). This is the proportion of breast tissue that contains the mass relative to the proportion of total tissue; thus it represents the average number of possible random locations radiologists could select, taking into account the lesion and image size across all of the target-present images. Figure 7a shows the percentage of trials when the radiologists responded correctly on localisation task, when detection was correct, for low density (blue line) and high density (black line) at the two durations, compared with chance. Single sample t-tests (Bonferonni adjusted, alpha = .0125) showed that radiologists' localisation accuracy was significantly above chance (4.4%) for 250ms presentations of low density images ($t(11) = 12.9$, $p < .0001$, $M_{diff} = 30.18$, CI [25.03, 35.33], $d_{unb} = 4.9$) as well as for high density images ($t(11) = 3.74$, $p = .003$, $M_{diff} = 6.43$, CI [2.64, 10.22], $d_{unb} = 1.42$). The same pattern was evident at the longer duration of 1000ms for low ($t(11) = 13.9$, $p < .0001$, $M_{diff} = 50.6$, CI [42.59, 58.61], $d_{unb} = 5.28$) and high ($t(11) = 10.41$, $p < .0001$, $M_{diff} = 19.35$, CI [15.26, 23.44], $d_{unb} = 3.95$) density images.

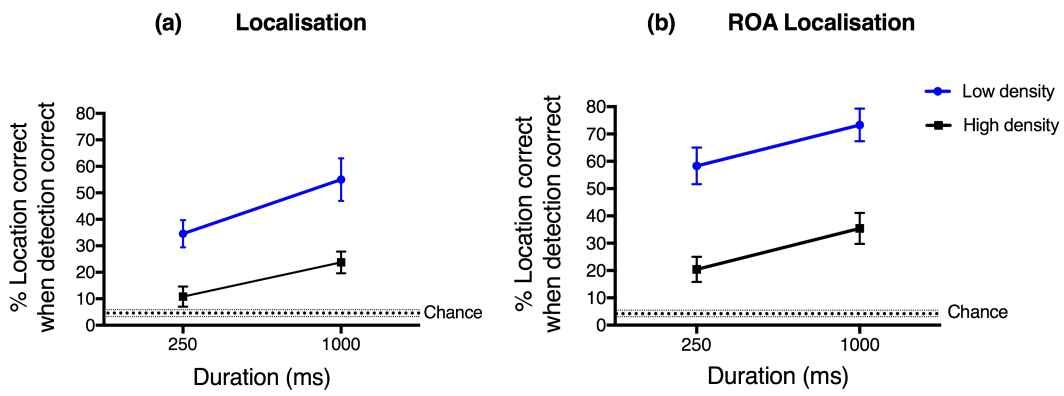


Figure 7: Detection and localisation results. (a) Average percentage correct on the localisation task for trials when detection was correct; (b) Average percentage correct on the localisation task when a region of acceptance (ROA) around the lesion is included. Chance is 4.4% (dotted line) with 95% confidence intervals. Error bars represent 95% confidence intervals.

To investigate the effect of density on localisation (Fig. 7a), we conducted a repeated measures ANOVA with the factors of Density (low, high) x Duration (250, 1000) on the mean percentage localisation correct values from the correct detection target-present trials. Again in line with expectations, this showed a main effect of Density, with better localisation accuracy in the low than high density condition ($F(1, 11) = 114.07$, $p < .0001$, $\eta^2_p = .91$), a main effect for Duration, with better localisation accuracy at 1000ms than 250ms ($F(1,11) = 53.01$, $p < .0001$, $\eta^2_p = .83$), and no Density x Duration interaction ($F(1,11) = 2.17$, $p = .17$). These analyses show that radiologists were statistically above chance in localising the target on trials where they successfully detected a mass. However, as localisation performance is far from perfect, we have some trials on which detection occurred apparently without localisation information being

available. This could reflect a global signal as suggested in the previous literature (Evans et al., 2013) and to investigate this possibility thoroughly, we conducted several follow-up analyses.

Before concluding one has evidence of 'detection without localisation' (e.g., Evans et al., 2013; Evans et al., 2016), there are some important alternatives to be considered. First, we would like to note that before concluding anything from a null localisation effect (such as that of Evans and colleagues), we need to use statistics that can provide evidence of *no effect* (of localisation when there is detection) rather than just no evidence. Frequentist statistics do not allow for the interpretation of null effects – a *p* value greater than alpha merely informs us that we do not have evidence to reject the null hypothesis. To see whether there is evidence *for* the null hypothesis of *no localisation* information, we could instead calculate a Bayes Factor (BF). In line with Jeffreys (1961) a BF < .3 indicates that the data support the null rather than the alternative hypothesis, a BF ~ 1 indicates maximal insensitivity of the experimental evidence, whereas a BF >1 indicates the data support the alternative hypothesis (BF > 3 suggests evidence for the alternative) (Dienes, 2011). In our case, we do not have a null effect in any condition, but we can still calculate a Bayes equivalent of a single sample t-test compared to chance (4.4%) to illustrate the point: if we test just the difficult images that are comparable to those of Evans et al. (2013; 2016), we can see strong evidence for the alternative hypothesis that localisation information exists: For the high density condition at 250ms, the BF(12) = 14.73 and at the longer duration, 1000ms, BF(12) = 31052.09. Consistent with our frequentist statistics results, we conclude that the radiologists are localising targets better than chance in the high dense conditions.

Our second consideration is whether summary level statistics such as overall accuracy or sensitivity are adequate to address the 'detection without localisation'

question. In fact, one cannot be sure of 'detection without localisation' without examining the error trials carefully. A null localisation effect could, for example, be due to less precision in the localisation task than the detection task, due to the additional requirements rather than a true lack of localisation information. This could include decay in the visual short-term memory trace over time or motor error in clicking the precise location. If such factors influence the precision of the localisation responses, we should see localisation errors that nonetheless cluster around the correct region. Our radiologists were scored correct on localisation if the mouse-click occurred within or on the boundaries of the lesion, consistent with Evans et al. (2016) (Evans, personal communication, May, 2017). However, when we look at the incorrect localisation responses, we see that this does not accurately reflect the degree of localisation information. For example, in Figure 8a, many of the 'incorrect' responses suggest the participant had some information about location, rather than basing his or her response on an amorphous global signal of abnormality.

There is also inherent variability in real-world stimuli. Although we carefully selected images with only one true mass, and removed obvious artefacts (e.g., dust), the images have naturally-occurring variations in breast tissue. We need to examine the responses at an image level to assess whether such variance may have contributed to trials of apparent successful detection without accurate localisation. Figure 8b shows clearly an image where natural variability has contributed to three incorrect responses to a distractor in the breast (presumably in these cases, the radiologists were responding 'abnormality present' to this distractor, rather than the actual mass). The responses on these images suggest that apparent 'detection without localisation' may actually reflect coarse or less precise localisation, rather than no localisation, warranting image-level investigation.

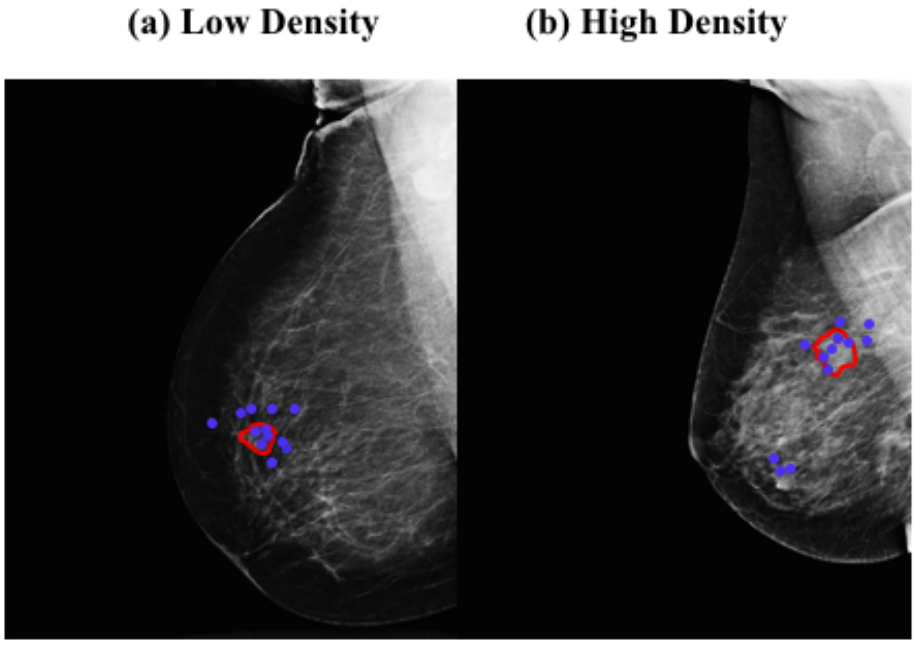


Figure 8: Exemplars from the target present stimuli set illustrating the mass (red outline, not shown in the experiment) and localisation responses of the 12 radiologists (blue) collapsed across duration. (a) Low density image showing precision errors. The blue mouse-clicks for localisation show that the 8 radiologists who were 'incorrect' on this image may have information about the location of the target; (b) High density image showing the effect of a naturally-occurring distractor. Three radiologists localised the distractor as the abnormality (note a further 4 'incorrect' responses are near the mass (red outline) but imprecise).

To quantify the degree to which such examples might influence our results, we conducted a post-hoc image analysis collapsed across participants for each duration. We calculated the distance between the response click and the mass (i.e., the degree of incorrect localisation). In academic radiology, a region of acceptance (ROA) for lesion localisation is determined by taking into account the size of the largest lesion (e.g., Haygood, et al., 2014). Following this convention, we measured the radius of the largest mass in the image set (27mm) and added this value to the boundary values for all the target present images. Using this method, localisation is scored correct when a radiologist clicks within this ROA, allowing for a margin of response imprecision and reducing the 'tightness' of acceptance. We further examined the trials that were still incorrect to quantify the distance from the lesion boundary.

Figure 9 shows image level analysis for the localisation data on incorrect trials plotted as a function of distance (in pixels) from the closest boundary of the mass, collapsed across radiologists (Fig. 9a: 250ms; Fig. 9b: 1000ms). Detection incorrect images are not included (250ms: high density =12, low density = 1; 1000ms: high density = 8; low density = 0). Correct responses for localisation (when detection correct) would appear on the baseline and are also not included on the figure (250ms: high density = 3, low density = 8; 1000ms: high density = 8; low density = 8). The dashed red line represents the ROA plotted at 29 pixels. Figure 9 shows a considerable proportion of the clicks lie within this decision boundary and highlights how the variability within each image affected accuracy due to factors such as mass size and distractors.

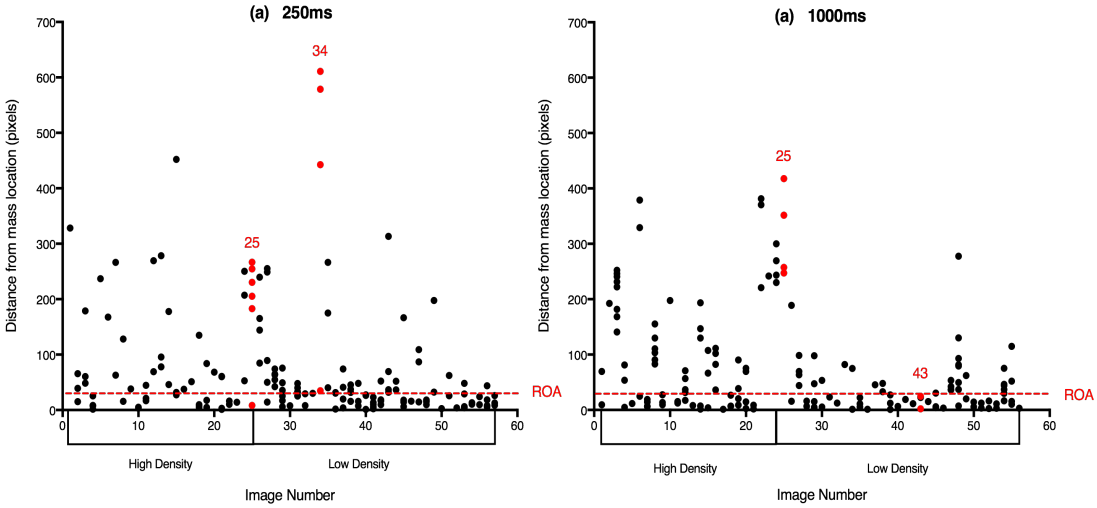


Figure 9: Localisation errors showing the distance between the localisation response and the mass for each image (detection correct target-present trials only). (a) 250ms duration; (b) 1000ms duration. The x-axis represents the images (divided by high and low density. Note: the image numbers are arbitrary for the purpose of the graph only). A correct score on localisation would score 0 (excluded from the figure). The y-axis is the distance (in pixels) from the mass border. The dashed red line represents the region of acceptance (ROA). Red numbers are data points in response to images with unusual characteristics: 25 (250ms) is the high density image presented in Figure 8b showing the mouse-clicks on a distractor. 34 is a low density image which contained a prominent lymph node in the axillary tail of the breast which appears to have captured 4 radiologists' attention; 25 (1000ms) is a low density image containing a small mass and 43 is the low density image presented in Figure 8a showing the cluster of mouse-clicks near the correct location.

Localisation accuracy including a ROA: We calculated percent correct for localisation trials with an ROA included in assessing localisation for target-present trials with correct detection responses. Figure 7b shows the percentage of trials in which ROA localisation was correct for low density (blue line) and high density (black line) images across both durations, compared with chance. The summary-level measures clearly indicate better accuracy for all conditions compared with the non-ROA data, especially for the 250ms high density condition (ROA Mean = 20.42%; non-ROA Mean = 10.83%), demonstrating that the Evans et al. (2013) and Evans et al. (2016) method for calculating localisation may not adequately capture the degree to which location information is present.

This post-hoc analysis highlights the variability and challenges which exist when using real-world stimuli, and the importance of carefully examining the data from individual images rather than stopping at summary statistics. These findings suggest that the apparent lack of localisation on some trials where a mass was detected is, at least in part, driven by image variability, such as small masses and distractors, and response imprecision. When we apply a more liberal localisation ROA, we see evidence that coarse localisation information exists, with a higher proportion of correct localisation responses even for the more difficult images.

We can also bin trials on which detection was correct according to their response profile to further examine the distribution of trial performance. Figure 10 shows the localisation data calculated using an ROA as a function of detection performance (collapsed across radiologists and images) for trials on which detection plus localisation were correct (blue bar), the additional localisation correct trials produced by including a

ROA (dark grey bar) and detection only trials on which localisation was incorrect (light grey bar).

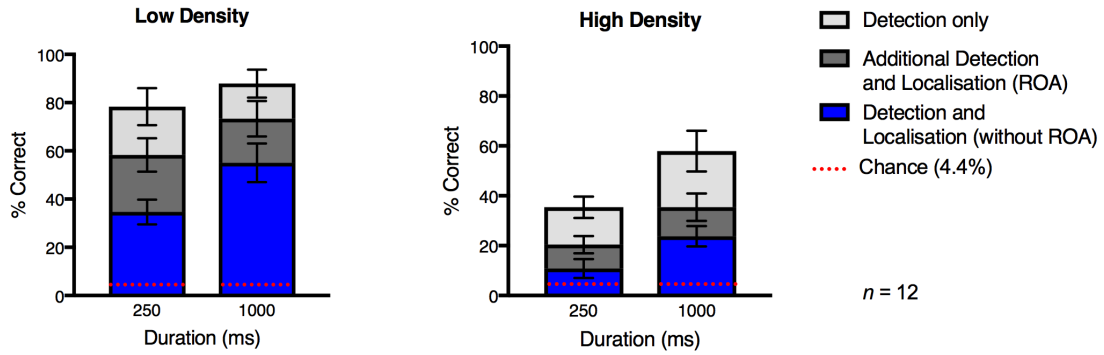


Figure 10: Percentage correct detection and localisation on target-present trials for low and high density mammograms plotted by duration (250ms, 1000ms). Data are separated by response accuracy: Detection and localisation correct (blue bar); the additional proportion of trials where localisation is correct when a ROA is included (dark grey bar); and detection correct/localisation incorrect (light grey bar). Error bars represent 95% confidence intervals.

In addition to the trials with evidence for coarse localisation or precise mis-localisation, Figure 8 shows some remaining trials on which localisation is clearly incorrect; these contribute to the light grey bars in Figure 10. These trials could be evidence for 'detection without localisation', which seems key to interpretations of radiologists using 'gist' or a global signal. However, there is one final consideration before making such an interpretation: we need to be sure that the number of trials on which this occurs exceeds the rate at which such trials would occur simply from 'lucky' guesses. With any visual detection task, some proportion of trials will be correct by

chance. A *d′* above chance shows more trials are correct than would be predicted by simply guessing, but if one wants to infer that there are trials in which there is 'detection without localisation', we need to calculate what proportion of these could be lucky correct detection guesses, followed by a localisation guess (which has less chance of being correct, recall chance in Evans et al. (2016) for localisation was ~ 6%).

We calculated a guessing probability using the method described in Howe and Webb (2014). They were interested in whether observers could ever 'sense' a change in a change blindness paradigm without knowing where the change was. In their method, one works out what proportion of correct detection trials (in their study, detection of a change) could be due to lucky guesses by creating a hypothetical observer who can only detect a change when it also knows what that change is (i.e., there is no true detection without localisation, therefore any such trials are due to correct guesses). Here, we used the same logic, a hypothetical observer who cannot detect a mass without also knowing where that mass is, to work out the proportion of trials on which correct detection combined with incorrect localisation could be due to lucky guesses. We can then compare actual performance with this prediction for each radiologist.

$$\textit{Calculated N (hypothetical observer)} = Q(Y-PA)/(1-P)$$

where Q = proportion of possible incorrect localisations; Y = number of target present trials on which the participant responded 'target present' (hits); P = proportion of target absent trials on which the participant responded 'target present' (false alarms); and A = actual number of target present trials (Note, there is no correction applied to an observer with no false alarms).

We calculate a guessing probability for the ROA localisation data, as this already takes into account any slight imprecisions in the localisation responses, giving the most accurate view of localisation information at a summary level. If the actual participants correctly indicated the presence of a mass in the absence of a correct location response more often than this hypothetical observer, this provides evidence for information about the presence of an abnormality without knowing where it is: 'detection without localisation'. Figure 11 shows the number of 'detection without localisation' trials from our data (dark grey bars) and the number of trials the hypothetical observer would 'guess' for all four conditions (light grey bars).

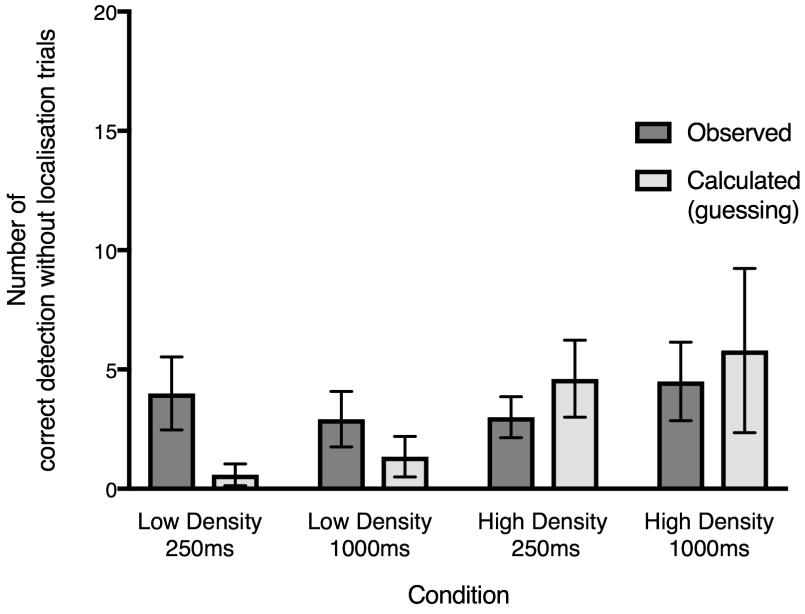


Figure 11: The observed number of correct 'detection without localisation' trials (dark grey bars) compared to the number of calculated (guessing) trials for a hypothetical ideal observer (light grey bars) for low and high density mammograms plotted by duration (250ms, 1000ms). Error bars represent 95% confidence intervals.

From Figure 11, it is clear that there are only a small number of trials representing apparent 'detection without localisation', which makes statistical analysis unlikely to be reliable. However, even just from the graph one can see that only for the low density conditions is there any chance that there might be more detection without localisation trials than predicted by our hypothetical observer. Recall that it is only our high density condition that has images in which the mass is comparable in difficulty to Evans and colleagues (2013; 2016), making this the key condition. We have no evidence that for this high density condition the number of observed 'detection without localisation' trials is more than what would be predicted by 'lucky' guesses.

Discussion

The aim of this study was to examine the type of information that is available in the initial processing of a medical image (mammogram) by experienced radiologists, focusing on *detection* and *localisation* of potential abnormalities. We found radiologists were able to *detect* abnormalities at both durations (250ms, 1000ms) and density conditions (high, low), with a significant effect of duration. Overall summary statistics also supported the presence of *localisation* information, with the radiologists performing better than chance for both the 250ms and 1000ms durations, for the low and high density mammograms. Breast density affected performance in a predictable way, with better performance for low than high density images. As our key question related to a potential dissociation between detection and localisation, we carefully examined trials on which there seemed to be a dissociation. We suggest a number of factors that can lead to an underestimation of localisation information such as image variability, the precision of localisation responses, and correct detection guesses. Overall, our data suggest that although it is possible that there may be a dissociation between detection and localisation

on a small number of trials, particularly on easy trials (low density), there are other plausible explanations for the majority of such apparent dissociation trials.

Recent high-profile papers have concluded that radiologists can detect but not localise abnormalities in briefly presented mammograms (Evans et al., 2013; Evans et al., 2016). These papers suggest a different process to the previous theory that the information in the first glance guides experienced radiologists' attention and directs their eyes towards the location of the potential abnormality (Kundel & Nodine, 1975). Specifically, Evans et al. (2013) and Evans et al. (2016) proposed that the information extracted in the early signal is a global impression, which alerts the radiologist to the presence of an abnormality and then prompts a more thorough search, rather than guiding attention to the region of the abnormality directly. This alternative theory depends crucially on radiologists being able to detect masses in the *absence* of any information about location.

One of the key distinctions between the Evans et al. (2013; 2016) studies and our study is that they presented stimuli described as 'subtle masses and architectural distortions'. This might mean that there were a mix of both potentially localisable abnormalities (subtle masses) and abnormalities that do not have a well-defined location (architectural distortions, which do not contain a discrete mass in the parenchyma), or each category in a separate image. It would then make sense if there were no localisation signals if the abnormality was not well defined in location. A global or gist signal also seems a plausible explanation for the other intriguing findings of this group in which radiologists are above chance in detecting abnormality in a patient when shown whole mammograms of a contralateral normal breast (Experiment 2) or only a patch of a mammographic image that does not actually contain the mass (Evans et al., 2016, Experiment 4). In these cases, however, there is no mass to localise, making these

findings less relevant to the question of whether a localisable mass can indeed be detected without being localised but may explain how a global signal can be used to diagnose an abnormality. The evidence pertaining to this question comes from the experiments in which it seems there is a mix of pathology (architectural distortions and subtle masses). It would therefore be interesting to know the proportion of these two types of breast pathology in the Evans et al. (2016) stimulus set, and how the location data break down by pathology. This would then allow a more accurate comparison with our own data.

There is also a potential concern with the analyses in the Evans et al. (2013; 2016) studies. Working with real-world images introduces many challenges, and with *d* prime values quite close to chance, we raise the concern that these data could be driven by a small number of images that contained additional artefacts. If these studies had any images like those illustrated in Figure 8, this could contribute to correct detection but incorrect location responses. Similarly, if the localisation responses cluster around the actual mass but not within the boundaries in some images, such as we found in our data (see Figure 9), this would also contribute to apparent detection without localisation. In the Evans et al. (2016) patch and contralateral breast experiments (2 and 4), coarse localisation cannot be an explanation, as there is no actual mass to localise. Thus, if there are any artefacts in images that drive detection above chance, this will appear to be dissociated from location (which is always chance). With only summary statistics showing *d* prime slightly (but significantly) above chance (Evans et al. (2013); Evans et al. (2016)), it is possible that the data interpreted as evidence for a global signal could be misleading.

Even when we use a conservative measure of localisation (click within the mass boundary), we were not able to replicate the findings of Evans et al. (2013) and Evans et al. (2016) that there are circumstances where radiologists can detect a mass above chance

but not localise it. This could simply reflect that we were not at exactly the right durations to catch a dissociation due to variability in the experience of the participants, difficulty of the images, and other cross-experiment differences between our study and those of Evans and colleagues (2013; 2016). Another potential factor that could influence the difference between the studies is that our participants seem to be more experienced than those of Evans and colleagues (2013; 2016). This may be reason that we found localisation at a summary statistics level: our more experienced participants could extract information more rapidly and therefore processed the images in greater detail. For these previous studies to make the inference that there is *no localisation,* however, still requires an approach such as Bayes statistics, rather than standard frequentist statistics. Here, we have outlined the steps that seem crucial to be able to make an inference of dissociation between detection and localisation.

Although at the summary statistic level we did not replicate the lack of localisation information, we did find trials on which detection responses were correct but those for localisation were incorrect. We were therefore able to use these to investigate factors that might contribute to an apparent dissociation between detection and localisation. First, variability in the target-present images might be contributing misleading data to the summary statistics. Using real-world stimuli rather than typical laboratory visual search displays allows for high ecological validity, but presents challenges. The available images are highly variable and it is difficult to control for factors such as co-existing variables (e.g., breast calcifications, target number and size and breast tissue type). Indeed, we identified images where there were clear clusters of incorrect localisation corresponding to a specific visual feature in the image, suggesting the detection response was based on an incorrect identification (i.e., of the distracting feature). Second, we find evidence that coarse localisation information is often present in

apparently incorrect responses. When we use a region of acceptance around the lesion, we see clusters of correct localisation responses surrounding the lesion. This suggests that task demands, such as having to hold the information through a detection response and subsequent location screen, may result in a loss of precision. Alternatively, it may be that the location information is only present at a coarse level in the first place (and is perfectly maintained). Finally, on trials where there is detection but incorrect localisation (by whatever definition one uses), it is important to consider the contribution of correct detection guesses. We used a method for estimating the effect correct guesses might have on the subsequent results. The key high density condition, which is most similar to that of Evans and colleagues (2013; 2016), gives no evidence for there being more 'detection without localisation' trials than can would be predicted to be lucky guesses. Thus, the pattern taken from a small number of trials suggest that in the difficult images, such as our set of high density mammograms, apparent 'detection without localisation' responses can be accounted for by 'lucky' guesses.

Our only evidence of an apparent dissociation between detection and localisation comes from the low density conditions. Intuitively, a salient mass seems most likely to have localisation information recorded, as there is a stronger bottom-up signal (much like a classic 'feature search'). Indeed, we do see overall better performance in the low density conditions compared with the high density conditions (although nowhere near 'pop-out' performance). Although our ROA takes into account coarse localisation information, it cannot account for image-level variability where a distractor may have been selected, or the potential decay of localisation information over time. Thus, while it is possible that these potential 'detection without localisation' trials in the low density condition could reflect a global signal that is used to make a detection response, as proposed by Evans et al. (2013) and Evans et al. (2016), these trials could alternatively

reflect the contribution of other factors to reducing localisation accuracy. Overall, such 'detection without localisation' occurred on a very small number of trials (~ 4), precluding statistical analysis, which means we have only the numerical difference to support any such inference. This means that for most of our stimuli, including those most similar to the previous studies, when the radiologists reported detecting a mass, they also had some information about where it was.

The proposal by Evans et al. (2013) and Evans et al. (2016) that radiologists use a global signal lacking in location information has important theoretical implications, as it identifies a very different mechanism from the Kundel and Nodine (1975) classic theory. Our results, however, demonstrate that successful detection of a mass in briefly presented mammograms is typically accompanied by information about location. This is more consistent with the Kundel and Nodine (1975) model: that the initial signal guides attention and eye movements to the lesion. To fully reconcile these distinctions, we need a study which investigates the presence (or lack thereof) of both global and localisable signals across three clearly defined conditions with different degrees of potential localisation (a salient mass, a subtle mass, or diffuse parenchymal change). We then need to ensure that the analyses are appropriate to the key question of whether any localisation information exists through a thorough image-level analysis.

Both detection and localisation performance decreased with increased breast density at fast presentations. These results are related to what we know about clutter in natural scenes and visual search in free viewing: increasing clutter or set size decreases performance (Adamo, Cain & Mitroff, 2015; Asher, Tolhurst, Troscianko, & Gilchrist, 2013; Rosenholtz, Li, Mansfield & Jin, 2005; Rosenholtz, Li & Nakano, 2007; Whitney & Levi, 2011; Wolfe, 1994). Fibroglandular tissue, which increases density on a mammogram, appears more radio-opaque than fat and may increase crowding and/or

masking effects reducing performance in the denser mammograms. In the medical perception literature, there have been a number of studies that have investigated factors such as lesion subtlety, which may be dependent on the surrounding anatomical structures (e.g., Krupinski, 2005). Analogous to clutter interfering with performance in natural scenes, our results show similar effects in radiologists interpreting medical images.

These findings improve our understanding of how density can influence a radiologists' diagnostic decision and therefore have clinical relevance. Female breast tissue is highly variable with regards to mammographic breast density (MBD: Li et al., 2013) and high levels of breast density reduce radiologist sensitivity (see Al-Mousa, et al., 2014). It has been suggested that what radiologists perceive and thus report in the first second is critical (Mello-Thoms, 2009), that women with dense breasts make up almost a half of the population (Sprague et al., 2014), and that there is an increased risk of developing cancer in dense breasts (Boyd et al., 2010). Our results confirm that MBD has a negative impact on mass detection and localisation when radiologists are shown an image briefly. From a clinical viewpoint, we should inform women and their clinicians about their MBD levels, for appropriate and personalised care. For instance, in the case of a dense breast, further imaging modalities such as 3D mammography (digital breast tomosynthesis), ultrasound or magnetic resonance imaging will facilitate a definitive diagnosis. Although for almost half of the United States, density scoring is included (Slanetz, Freer & Birdwell, 2015), current breast screening reporting protocols in Australia do not include a mammographic density rating. Our data shows that high breast density reduces the amount of information available in the first glance, suggesting reporting this information should be mandatory.

Conclusions

Here, we explored the degree to which information available in very brief presentations of medical images can support both detection and localisation of a mass in mammograms. Access to location information is crucial for guiding actions or further analysis (e.g., eye movements). We find a tight link between information supporting detection and localisation, and present methods that allow a stronger test of the claim that detection of a mass can occur based on gist (without knowledge of location). Although it is certainly possible that gist and the non-selective pathway of visual processing contribute to the detection of a non-localisable abnormality, our systematic examination of the factors that can result in apparent dissociation between detection and localisation demonstrates the importance of going beyond summary statistics when seeking to test this hypothesis. We emphasise the importance of considering factors such as stimulus variability, response imprecision, and participant guessing. Our results are consistent with Kundel & Nodine's (1975) model of radiologist visual search suggesting that the initial signal in a brief glance contains information that subsequently guides attention to the abnormality. Finally, we suggest the finding of reduced performance for dense mammograms illustrates the importance of reporting density information in the context of medical screening.

References

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2015). Targets Need Their Own Personal Space: Effects of Clutter on Multiple-Target Search Accuracy. *Perception*, *44*(10), 1203–1214. http://doi.org/10.1177/0301006615594921.

Asher, M.F., Tolhurst, D.J., Troscianko, T., & Gilchrist, I.D. (2013). Regional effects of clutter on human target detection performance. *Journal of Vision*, *13*(5)*:25*, 1-15.

Al Mousa, D. A., Ryan, E. A., Mello-Thoms, C., & Brennan, P. C. (2014). What effect does mammographic breast density have on lesion detection in digital mammography? *Clinical radiology*, *69*(4), 333-341.

American College of Radiology (2013). Breast Imaging Reporting and Data System (BI-RADS) Atlas. Reston, VA, American College of Radiology.

Berbaum, K. S., Krupinski, E. A., Schartz, K. M., Caldwell, R. T., Madsen, M. T., Hur, S., Laroia, A.T., Thompson, B.H., Mullan, B.F., & Franken, E. A. (2015). Satisfaction of search in chest radiography 2015. *Academic radiology*, *22*(11), 1457-1465.

Boyd, N. F, Martin, L. J, Bronskill, M, Yaffe, M. J, Duric, N & Minkin S. (2010). Breast tissue composition and susceptibility to breast cancer. *J Natl Cancer Inst, 102*(16), 1224-1237.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. doi:10.1163/156856897X00357

Buswell, G. T. (1935). How people look at pictures: A study of the psychology of perception in art. *Chicago: University of Chicago Press.*

Bulu, H., Alpkocak, A., Balci, P. (2013). Uncertainty modelling for ontology-based mammography annotation with intelligent BI-RADS scoring. *Computers in Biology and Medicine, 43*(4), .301-311. DOI: 10.1016/j.compbiomed.2013.01.001.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*(3), 274-290.

D'Orsi, C. J., Getty, D. J., Pickett, R. M., Sechopoulos, I., Newell, M. S., Gundry, K. R., Bates, S.R., Nishikawa, R.M., Sickles, E.A., Karellas, A., & D'Orsi, E. M. (2013). Stereoscopic digital mammography: improved specificity and reduced rate of recall in a prospective clinical trial. *Radiology*, *266*(1), 81-88.

Evans, K.K., Georgian-Smith, D., Tambouret, R., Birdwell, R.L., & Wolfe, J.M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin Review, 20*(6)*,* 1170 – 1175. doi: 10.3758/s13423-013-0459-3

Evans, K. K., Haygood, T. M., Cooper, J., Culpan, A.-M., & Wolfe, J. M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences*, *113*(37), 10292–10297. http://doi.org/10.1073/pnas.1606187113.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137-176.

Haygood, T. M., Ryan, J., Brennan, P. C., Li, S., Marom, E. M., McEntee, M. F., Itani, M., Evanoff, M., & Chakraborty, D. (2014). On the choice of acceptance radius in free-response observer performance studies. *The British journal of radiology, 86*(1021), pp.42313554-42313554.

Howe, P.D.L. & Webb, M.E. (2014). Detecting unidentified changes. *PloS ONE, 9*(11), e84490.

Jeffreys, H. (1961). *The Theory of Probability* (3rd ed.). Oxford, England: Oxford University Press.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorisation and object interference. *Vision Research, 47*(26), 3286 – 3297.

Knutzen, A. M., & Gisvold, J. J. (1993). Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions. In *Mayo Clinic Proceedings*, *68*(5), 454-460.

Krupinski, E. A. (2005). Visual search of mammographic images: Influence of lesion subtlety. *Academic radiology*, *12*(8), 965-969.

Kundel, H. L., & La Follette Jr, P. S. (1972). Visual search patterns and experience with radiological images 1. *Radiology*, *103*(3), 523-528.

Kundel, H.L., & Nodine, C.F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*(3), 527-532.

Kundel, H. L., Nodine, C. F., Krupinski, E. A., & Mello-Thoms, C. (2008). Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic radiology*, *15*(7), 881-886.

Li, J., Humphreys, K., Eriksson, L., Edgnen, G., Czene, K., & Hall, P. (2013). Mammographic density reduction is a prognostic marker of response to adjuvant tamoxifen therapy in postmenopausal patients with breast cancer. *Journal of Clinical Oncology, 31*(18), 2249-2256.

Mello-Thoms, C. (2009). The "Holistic Grail": Possible implications of an initial mistake in the reading of digital mammograms. *Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment (Proceedings of SPIE, Vol. 72630)* doi: 10.1117/12.813778

Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., Stalder, S., & Maitz, G. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Academic Radiology, 12*(7)*,* 830-840.

Nodine, C.F., Kundel, H.L., Mello-Thoms, C., Weinstein, S.P., Orel, S.G., Sullivan, D.C., & Conant, E.F. (1999). How experience and training influence mammographic expertise. *Academic Radiology, 6*(10)*,* 575-585.

Nodine, C.F., & Mello-Thoms, C. (2010). *The handbook of medical image perception and techniques* (pp. 139-156). E. Samei, & E. A. Krupinski (Eds.). Cambridge: Cambridge University Press.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem. 2*(5), 509–522.

Rawashdeh, M. A., Lee, W. B., Bourne, R. M., Ryan, E. A., Pietrzyk, M. W., Reed, W.M., Heard, R.C., Black, D.A. & Brennan, P.C. (2013). Markers of good performance in mammography depend on number of Annual readings. *Radiology, 269*, 61 – 67.

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005, April). Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 761-770). ACM.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17–17. http://doi.org/10.1167/7.2.17.

Slanetz, P. J., Freer, P. E, & Birdwell, R. L. (2015). Breast-density legislation - practical considerations. *New England Journal of Medicine, 372*(7), 593-595.

Sprague, B. L., Gangnon, R. E., Burt, V., Trentham-Dietz, A., Hampton, J. M., Wellman, R. D., Kerlikowske, K., & Miglioretti, D. L. (2014). Prevalence of mammographically dense breasts in the United States. *Journal of the National Cancer Institute*, *106*(10), dju255

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 4–4. http://doi.org/10.1167/7.14.4

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 382*(6582), 520 – 522.

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artefactual objects. *Perception, 30*(6), 655-688.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, *15*(4), 160-168.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in cognitive sciences*, *15*(2), 77-84.

Chapter 4: The influence of prior expectation and expertise on attentional cueing in medical images

Chapter summary

Radiologists make critical decisions based on searching and interpreting medical images. Prior expectations may set a search strategy or attentional bias, as the probability of a nodule differs across anatomical regions within the tissue. Using a modified attention-cueing paradigm, we investigated the potential for information in medical images to cause attention shifts in naïve participants and radiologists. Our first aim was to test whether priors about likely locations for nodules in chest radiographs affected the allocation of attention. Our second aim was to see whether expertise boosted the salience of subtle signals, resulting in attentional capture by nodules in radiologists that did not affect attention in naïve participants. Both groups started with a block of 'no prime' target detection trials to assess any underlying bias across the vertical meridian. In block two, normal chest radiographs were presented as primes before a target dot to test the effect of priors alone. In block three, chest radiographs with a single pulmonary nodule (equally distributed between the left and right lung across images) were presented as primes. For the naïve participants, we presented an additional block where the nodule was artificially enhanced to give a strong 'bottom-up' signal. For the radiologists, we only presented the original subtle nodule images. The task was a simple visual detection, responding as quickly as possible to the appearance of a small dot that appeared following the prime display. In naïve observers, the *salient* nodule (but not the original *subtle* nodule) cued attention, validating our paradigm. For the radiologists, a Bayes analysis revealed there was inclusive evidence that attention shifted to locations based on statistical probability of a nodule, and no evidence that the nodules did not cue attention. There were, however, some intriguing hints that prime information (including nodule location) is available after brief durations for the more experienced radiologists. Experience may lead to higher

sensitivity to the *subtle* nodules, in that there was a positive correlation between more experience and the validity effect.

In this chapter, I look for evidence that priors, based on experience about the statistical probability of nodule locations, cause a shift in attention, and the extent to which information in a medical image can cue spatial attention. In Chapter 3, a direct measure was used to see whether radiologists had information about the location of a mass after seeing an image briefly. Here, this is examined from a different perspective, exploring indirectly the consequences of expertise and information about mass location on the deployment of attention.

Introduction

There is evidence that certain pathologies occur more frequently in specific locations. For example, breast cancer is more likely to occur in the upper, outer quadrant of the female breast than other regions, due to the higher proportion of breast tissue in this region (Lee, 2005). In the lung, primary malignant nodules are one and a half times more likely to occur in the right lung than the left, and mostly in the upper lobe (Garland, 1961; Swensen, et al., 2000; Winer-Muram, et al., 2002). Disease of the lung includes primary cancers as well as single pulmonary nodules (SPN). SPNs are frequently (0.2%) present in chest radiographs (Holin, Dwork, Glaser, Rikki & Stocklen, 1959). Although these nodules are often benign when less than 3cm in size, they are at risk of developing into malignant disease (Midthun, Swensen & Jett, 1993), making them a key target for radiologists during screening. As lung cancer is an insidious process (develops gradually), and has an overall mortality rate of 85% (Siegelman, et al., 1986), SPNs are clinically significant and would therefore be of importance in a screening environment.

Experienced chest radiologists are therefore exposed to SPNs, most frequently in the upper right lung, in their daily work.

Targets of a visual search in the real world, whether a mass in a mammogram or a weapon in a bag, sit in relation to surrounding objects that give them global context (Biederman, 1972). Laboratory based studies (using non-medical images) have shown that participants are sensitive to the global context and the statistical regularities in a display. This type of incidental statistical learning has been shown to affect the allocation of spatial attention within visual search displays, referred to as contextual cueing (Chun & Jiang, 1998; Jiang, Swallow, & Rosenbaum, 2013). In a series of experiments, Chun and Jiang (1998) presented different spatial layouts (global context) of objects (Ts) among distractors (Ls). The task was to discriminate the orientation of the target T. Half of the layouts were repeated across the experimental blocks in which the target object location remained constant. Participants were unaware of the repetition. The results showed that when the display was repeated, targets were discriminated more quickly, meaning that the context of the target was implicitly learned during the experiment. For radiologists, the global context of a target in a medical image may well form a similar type of contextual cue. In lung images, then, they may be biased towards the right upper lung because of experience with SPNs appearing in this location more frequently.

The potential influences on attention described here are top-down and implicit, akin to what has been previously described in the natural scene literature as 'scene based guidance' (Wolfe, Võ, Evans, & Greene, 2011). This guidance is thought to result from the build-up of a cognitive representation of how specific scenes appear (e.g., a kitchen), and how the items contained within it are spatially represented. It is influenced by scene structure and the sum of our past experiences. For those with experience reading medical

images, such ‘image based’ guidance, or priors, might similarly guide attention when one is within the context of the medical expertise.

One way to test for a shift of attention is to look at the effect of a cue or prime stimulus on performance of a subsequent task. In classic spatial orienting experiments (e.g., Posner, 1980), participants are asked to detect a visual target presented at a left or right peripheral location in a display. On each trial, a prime stimulus appears prior to the target display. In exogenous cueing paradigms, the prime appears in either the same location as the subsequent target (valid; 50% of trials) or in the opposite location (invalid; 50% of trials). The prime captures attention to its location, and this causes a measurable effect on the response to the target (a cueing effect: valid reaction time (RT) < invalid RT). In endogenous cueing paradigms, the prime is centrally presented and validly cues the subsequent target location on more than 50% of trials (e.g., 75% valid: 25% invalid). This results in a cueing effect driven by a voluntary shift of attention. These paradigms allow responses to an unrelated lateralised visual target to provide an index of the influence of a preceding prime stimulus on spatial attention.

Exogenous cueing depends on attention being captured by a salient stimulus. This is a ‘bottom-up’ effect, such that the observer cannot avoid having their attention drawn to the location of the prime (Jonides, 1981). In medical images, then, naïve observers may get capture from nodules that have high contrast relative to the rest of the prime image, but are unlikely to be captured by subtle SPNs. For radiologists, however, it is possible that their experience in detecting such stimuli could result in otherwise subtle nodules having a stronger signal. As seen in Chapter 3, radiologists can detect abnormalities in in brief displays (also in Evans, Georgian-Smith, Tambouret, Birdwell & Wolfe, 2013; Evans, Haygood, Cooper, Culpan & Wolfe, 2016; Kundel & Nodine, 1975). This might be due a higher sensitivity than naïve observers to certain abnormality-related signals,

which could capture attention to the location of the nodule in the prime. Thus, attentional cueing is a useful indirect measure of the extent to which location information is extracted from the prime display.

Nodine and Krupinski (1998) proposed that after long hours of perceptual learning, experts develop skills which are selectively tuned for relevant features when performing a task. For instance, experts in domains where performance maps onto recognising perceptual layouts in a display, such as master chess players, have demonstrated superior abilities when compared to novices. These experts have dedicated years of practice developing their skills and can detect the layout of a chess display in a glance (Charness, 2001). Radiologists with experience in breast mammography have been shown to fixate their eyes on 67% of cancers within one second of viewing a mammogram (Kundel, Nodine, Krupinski, Mello-Thoms, 2008). Those with experience reading chest radiographs are 70% accurate on detecting abnormalities after these images are flashed for 200ms (Kundel & Nodine, 1975). These experts are thought to recognise deviations from normal structures (or layout) rapidly, thus identifying abnormalities. The information extracted at this early stage is mostly based on pattern recognition and is compared with a cognitive template of 'normal' to reach a diagnostic decision (Nodine & Mell-Thoms, 2010). Indeed, in the vision literature, studies have shown that 'search templates' are set up, which direct attention to objects with shared features (e.g., Chun & Jiang, 1998). If radiologists have greater sensitivity to the features (or pattern) of a nodule, they should be more sensitive (i.e., 'tuned') to these features in medical images, or have an 'attentional set' for specific nodule features (Folk, Remington, Wright, 1994) compared with naïve observers. Therefore, radiologists should be 'tuned' to nodules in a chest radiograph more than naïve observers who have no prior experience reading medical images.

Here, I present the results of two experiments with naïve participants and radiologists which explored the allocation of attention using a novel modification of a Posner cueing paradigm (1980). Chest radiographs were presented as primes and the task was a simple visual detection of a lateralised (left or right) low contrast stimulus. For the first experiment, naïve observers saw four blocks of trials, completed in a set order. In the first block of trials, there was no prime, to test for any underlying attentional bias. In the second block, normal chest radiographs were presented as primes, to test whether simply invoking the context of the medical image would result in a bias in radiologists' attention towards more likely locations for nodules. In the third block, chest radiographs which contained a single, equally located (left and right lung) SPN, were presented as primes, followed by the same visual detection task. In the fourth block, these original chest radiographs were presented where the same nodule was artificially enhanced by creating a hybrid nodule, to give a strong 'bottom-up' signal. Here, we tested whether a salient cue in the context of a medical image causes a shift of attention to validate our paradigm. For the second experiment, radiologists performed the identical tasks but only performed three blocks: the no prime condition, normal chest radiographs, and the original nodule chest radiographs. To understand whether the observers had any explicit knowledge about the likely location of nodule, a brief questionnaire was completed at the completion of the experiment (See Appendix B).

If simply activating a context with a medical image activates priors regarding nodule likelihood, radiologists may show faster responses to subsequent targets appearing on the side most likely to show nodules (top right) than to targets appearing at other locations. In our nodule images, if salient contrast areas in cluttered images capture attention, naïve participants should get priming effects from the hybrid images. Thus, responses to targets on the same side as the enhanced nodule (valid) should be faster than

those on the opposite side (invalid). Similarly, if subtle cues are effective to capture a radiologist's attention due to their expertise with chest radiographs, they should show priming from a brief glance at a chest radiograph containing an original nodule. This would also act as an indirect method for seeing if localisation information is present from this brief exposure. We predict that there will be a priming effect where faster reaction times for the valid target locations relative to the invalid target locations occur. This could occur for the radiologists but not the naïve participants because of their tuning to relevant features of the nodule, and their ability to extract information about abnormalities from brief displays (including localisation information).

Experiment 1: Attentional cueing from medical images in naïve observers

Experiment 1 was designed to validate the medical image attentional cueing paradigm and find the appropriate experimental timing settings. In standard exogenous cueing, typical timings would be ~ 100ms prime-target delay (Posner, 1980). If one uses prime-target delays in excess of ~300ms, there is potential for a reversal of effects (invalid<valid), reflecting a phenomenon known as *Inhibition of Return* (IOR: Posner & Cohen, 1984) For endogenous cueing, prime-target delays typically need to be ~300ms or greater, with effects that persist over long time periods (valid<invalid). Here, we initially artificially boost the bottom-up signal of the nodule to create a salient cue within the medical image. This should therefore result in exogenous cueing. However, for radiologists to extract information about the context from the medical image, and therefore to allow exploration of potential effects of expertise, the prime image has to be on for a minimum of 200ms. We therefore needed to initially validate the paradigm to be

sure we could still get evidence exogenous cueing before examining the potential for expertise to affect perception of the primes and subsequent attentional allocation.

Method

All measures and conditions are reported.

Participants

A group of 23 naïve observers (4 male, mean age = 23 years, SD = 7 years) volunteered from the Macquarie University subject pool for course credit, after giving informed consent. All reported normal or corrected-to-normal vision and the study was approved by the Macquarie University Human Research Ethics Committee (Medical Sciences).

Stimuli and Apparatus

The central fixation point was a cross measuring 0.5° of visual angle which appeared against a grey background (RGB triplet: 200, 200, 200) and the target was a low-contrast grey circle (RGB triplet: 195, 195, 195; 1 degree in diameter). The prime consisted of 124 de-identified, posterior-anterior chest radiographs (62 normal, 62 SPN), downloaded from the Japanese Society of Radiological Technology database (JSRT: Shiraishi, Katsuragawa, Ikezoe, Matsumoto, Kobayashi et al., 2000), which is publically available at http://www.jsrt.or.jp/jsrt-db/eng.php . These images are digitized to 12 bits posterior–anterior chest radiographs with a resolution of 2048 × 2048 pixels (the size of one pixel is 0.175 × 0.175 mm^2). The nodule diameters range from 8 to 37mm (mean = 19 mm), they are located throughout the lungs (also behind the heart and under the diaphragm), and their intensities (densities) vary from nearly invisible to very bright. The nodules are subdivided in five categories, based on the degree of subtlety for

detection (rated by three chest radiologists), which is influenced by the nodule size, occlusion by other structures and nodule density. For the experimental trials, we selected images with a range of nodule subtlety: 3: subtle, n = 20; 4: relatively obvious, n = 17; and 5: obvious, n = 11. Half of the images contained a nodule in the left lung and half in the right. Nodule subtlety was balanced across location.

To validate the attentional cueing task, we modified the original SPN images to enhance the bottom-up salience of the nodule. These hybrid (salient) images were created using Adobe Photoshop (version CS6). As the lungs varied on factors such as size (e.g., degree of inspiration) and co-existing pathology (e.g., pneumonia), we wanted to present the same chests (true SPNs) that the radiologists would see, but boost the saliency of the nodule by the addition of a larger, higher contrast nodule, spatially superimposed over the original nodule. We selected the largest nodule (4.5° of visual angle) from the database and inserted it over all the SPNs in the original images. Figure 1a shows exemplars from the original (subtle) nodule SPN images, and Figure 1b from the hybrid (salient) nodule images. There were 48 images in each image set and nodule location was balanced across lung field (50% left, 50% right).

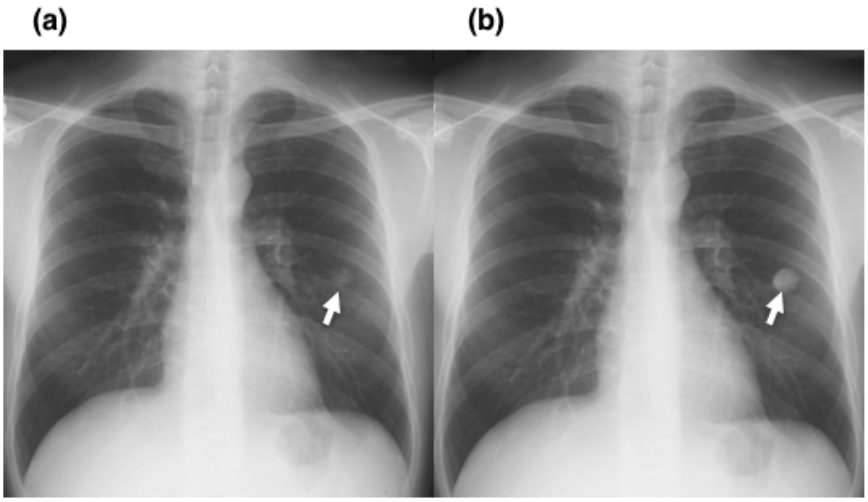


Figure 1: Exemplars from the target-present image sets. (a) Native nodule presented in the subtle nodule condition (white arrow); (b) Hybrid nodule presented in the salient nodule condition (white arrow).

Experimental sessions took place in a dimly-lit, windowless laboratory at Macquarie University, Sydney. Stimuli were downsized to 800 x 800 resolution and presented with MATLAB via PsychToolbox 3 (Brainard, 1997; Pelli, 1997), were flipped horizontally so that the right lung appeared on the left side of the screen and centred on a 1920 x 1080 resolution 27in Samsung SyncMaster AS950, refresh rate of 120Hz. The horizontal flip was to replicate the projection radiologists view in clinical practice. The observers sat at approximately 70cm away from the screen. The prime subtended approximately 20° of visual angle and the target subtended approximately .8° of visual angle.

Procedure

There were four experimental conditions, blocked and presented in a set order. In all blocks, the task was simply to press the spacebar as quickly as possible when the low-contrast target appeared. The target was displaced equally to the left or right from the centre by 240 pixels with a randomly varying radius of 120 pixels around this location.

The participants were asked to look at the cross and press the space bar as soon as a grey circle appeared. If no response, each trial timed out after 4 seconds. A new trial was indicated by a 100ms flash of the fixation cross (see Fig. 2). A simple detection task avoids high working memory or other tasks demands but it is important to ensure that the participant is not automatically pressing the response button. To reduce anticipatory responses, we added 12 catch trials where no target appeared. If the participants responded to a catch trial they received a red error message on the screen ('Error! No target'). In addition, we also randomly jittered target onset times relative to the prime offset (58, 92, 125ms). The Mean prime-target SOA was 342ms (specific durations: 308ms, 342ms and 375ms)

The first two conditions, *baseline* and *chest priors*, were composed of a practice block of 4 trials followed by 96 experimental trials, giving a total of 24 trials/condition (left/right). The *chest priors* condition included the context factor, which was a normal chest prime displayed for 250ms, after the fixation screen. Each chest prime appeared on two different trials (target left/target right) resulting in 96 experimental trials. In total, the participants saw 108 trials (including 12 catch trials) each for the *baseline* and *chest prior* conditions.

For the *subtle* and *salient* conditions, the normal chest primes were replaced with the relevant set for each condition (original nodule and enhanced nodule). As we have an additional factor (cue validity: valid/invalid) in the *subtle* and *salient* conditions, we repeated the primes in random order within each condition (observers saw each prime a total of 4 times: twice in the valid and twice in the invalid). Each condition was composed of 4 practice trials and 216 experimental trials, giving a total of 48 trials/condition. For each of these two conditions, we included 24 catch trials. In total, the participants saw 240 trials each for the *subtle* and *salient* nodule conditions. Breaks were given for 10

seconds every 48 trials and no feedback was provided. Note: we did not backward mask the stimuli so although timings were precisely programmed, the resulting processing time is only approximate.

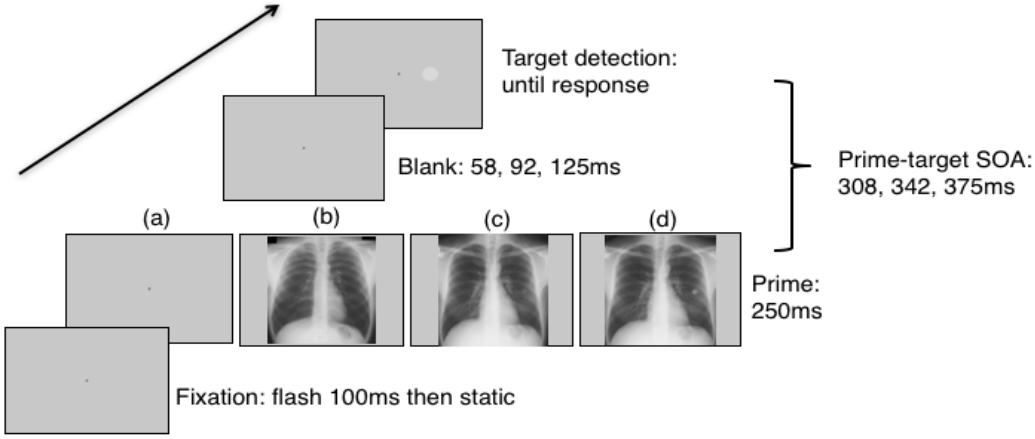


Figure 2: Example trial shown to 23 naïve participants in Experiment 1a. Trials began with a fixation cross. In separate blocks, the prime display was either (a), (b), (c) or (d) corresponding to our *baseline*, *chest priors*, *subtle* and *salient* conditions. The prime was displayed for 250ms followed by a variable blank screen with fixation for 58, 92 and 125ms (randomised within conditions) followed by the simple dot detection (note: target larger and brighter for illustration purposes). The prime-target SOA therefore varied between 308, 342 and 375ms (within each block)

Results and Discussion

Analysis

Mean differences (M_{diff}) with 95% confidence intervals (CI), as well as a Cohen's *d* estimate of effect size corrected for small sample size, to assist in accurate interpretation of the effects are presented. This latter measure, d_{unb}, represents an adjusted, unbiased Cohen's *d* standardised effect size applied to single sample t-tests where d_{unb} =

(1 - 3 / (4*df - 1)) * *d* (Cumming, 2012). For all analyses (including Bayes Factor) the Statistical Package for the Social Sciences was used (IBM: SPSS, 2015).

Our dependent variable for target detection was reaction time (RT, milliseconds). Outliers (defined as RTs less than 100ms and greater than 1000ms) were removed prior to statistical analyses. With these criteria, 1% of trials were discarded (across all conditions). Catch-trial errors ranged from 2-8% across all conditions and were removed from further analyses.

Baseline condition: When participants were asked to respond to the visual target with no prime, there was no difference in the RTs for left-sided targets and for right-sided targets. Figure 3 (left side bars) shows the mean RTs for the left-sided targets (blue bar) and the right-sided targets (grey bar). A paired samples t-test showed no significant difference [$t(22) = -.686$, $p = .5$, $M_{diff} = 2.91$, CI [-5.88, 11.69], $N = 23$, $r = .97$]. As in Chapter 3, we calculated a Bayes Factor (BF), which in this baseline condition provided some evidence for the null hypothesis, BF(23) = .27. We can conclude that the participants are not localising targets faster on the left or right side. We therefore have no evidence of an underlying bias.

Chest priors condition: When participants were asked to respond to the visual target following a normal radiograph prime, the RTs for the left-sided targets were different to the right-sided targets. Figure 3 (right side bars) shows the mean RTs for the left-sided targets (blue bar) and the right-sided targets (grey bar). A paired samples t-test showed that this difference in RT was significant: RTs were faster in the left compared to the right hemi-field [$t(22) = -3.22$, $p = .004$, $M_{diff} = 18.17$, CI [6.45, 29.89], $N = 23$, $r = .96$]. In standardised terms, this is a small effect ($d_{unb} = .18$, 95% CI [.059, .31]).

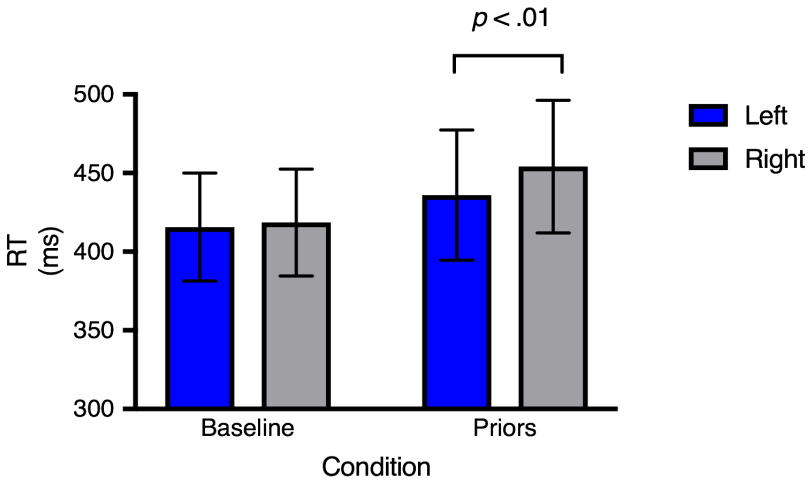


Figure 3: Experiment 1a. Mean Reaction Time (RT) for the naïve observers' performance on the cueing task for the baseline condition (left bars) and the chest priors condition (right bars). The blue bars represent the mean RTs for the left-sided target trials and the grey bars represent the mean RTs for the right-sided target trials. Error bars represent 95% confidence intervals.

To maintain ecological validity of the images, we did not alter the appearance of the radiographs. The significant cueing result in the priors condition (unexpected in naïve participants) may reflect effects of the stomach bubble (normal stomach filled with contents) which was present in some cases on the lower right of the screen (anatomical left). For participants who are naïve to reading chest radiographs, this normal anatomical variant could have caused a masking effect which slowed the participants down when the target appeared on the same side (right).

Subtle nodule condition: When participants were asked to respond to the visual target following a prime radiograph containing an original (*subtle*) nodule, the RTs for

targets appearing on the same side as the nodule (valid) were no different than for targets appearing on the opposite side as the nodule (invalid). Figure 4 (left side bars) shows the mean RTs for the valid targets (blue bar) and the invalid targets (grey bar). A paired samples t-test confirmed that this difference was not significant [$t(22) = 1.03$, $p = .32$, $M_{diff} = -2.37$, CI [-7.15, 2.41], $N = 23$, $r = .99$]. The Bayes factor was calculated as BF(23) = .35. This provides some support for the null hypothesis and indicates that the naïve participants do not show evidence for cueing effects for the *subtle* nodules.

S*alient nodule condition:* When participants were asked to respond to the visual target following a prime radiograph containing a *salient* nodule, the RTs for targets appearing on the same side as the nodule (valid) were slower than for targets appearing on the opposite side as the nodule (invalid). Figure 4 (right side bars) shows the mean RTs for the valid targets (blue bar) and the invalid targets (grey bar). A paired samples t-test showed that this difference was significant: unexpectedly, RT was faster in the invalid compared to the valid trials [$t(22) = 2.28$, $p = .033$, $M_{diff} = -8.68$, CI [-16.57, -0.78], $N = 23$, $r = .98$]. In standardised terms, this is a small effect ($d_{unb} = -0.11$, CI [-0.21, -0.01]). On review of the images, it appears that there is a masking effect such that when the target appeared on the same side as the *salient* nodule, it often was more difficult to see than when it appeared on the other side, which is reflected in the RTs.

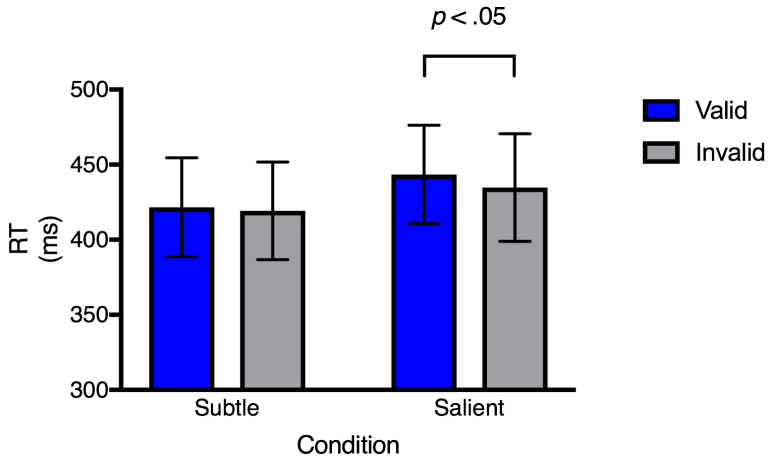


Figure 4: Experiment 1a. Mean Reaction Time (RT) for the naïve observers' performance on the cueing task for the subtle condition (left bars) and the salient condition (right bars). The blue bars represent the mean RTs for the valid target trials and the grey bars represent the mean RTs for the invalid target trials. Error bars represent 95% confidence intervals.

In summary, in Experiment 1a, we tested a group of naïve participants using a medical imaging attention cueing paradigm. There was no hemi-field bias for the *baseline* condition where only a simple visual detection task was performed. In the *chest prior* condition there was a small cueing effect (left < right), which may be due to the stomach (present in some of the images on the lower right side of the screen). The *subtle* prime failed to cue attention with support for the null hypothesis. Finally, in the *salient* prime condition, we expected to see an attention cueing effect in the valid trials compared with the invalid trials. However, there was a reverse effect, with RT faster for the invalid trials compared with the valid trials. This seems likely to be due to the strong bottom-up signal in the *salient nodule* condition masking the target in the subsequent display. Alternatively, this effect might reflect IOR (Posner & Cohen, 1984). We cannot, however, distinguish these possibilities in this experiment. We therefore repeated the

subtle and *salient* nodule conditions in a follow up experiment which avoided potential masking effects.

Experiment 1b

For an involuntary shift of attention to result in a validity effect, the experimental timings are crucial. In Experiment 1a, it is possible that the results are explained by IOR or it could be due to masking effects of the prime on the target. Experiment 1b was designed to reduce the potential masking effect of the nodule, and again test for an attention cueing effect from a chest radiograph prime in naïve observers. We presented only the *subtle* and *salient* conditions and modified the timing parameters, with the blank screen (either 50ms or 100ms in separate blocks) between the image (prime) and target (grey dot) adjusted to avoid masking effects.

Methods

All measures and conditions are reported.

Participants

A naïve group of 13 participants (male 1, mean age = 19.69 years, SD = 2.25 years) from the Macquarie University subject pool participated for course credit. Note: our sample size was selected *a priori*, to match the number of available radiologists for Experiment 2 (When conducting research on experts, the sample size is often constrained by the availability of the participants).

Stimuli and Apparatus

The apparatus, stimuli and design were unchanged except for the following modifications. As it has been reported that an experienced radiologist can detect an

abnormality in a chest radiograph after a presentation of 200ms (e.g., Kundel & Nodine, 1975), we reduced the prime duration from 250ms to 200ms. The observers saw primes and targets in two blocks (order counterbalanced) of (1) Mean SOA of 342ms (specific durations 308, 342, 375); (2) Mean SOA of 392 (specific durations 358, 392, 425 (randomly intermingled). These screen durations were chosen to capture an attentional cueing effect whilst trying to minimise the chance of IOR. In addition, to ensure the target appeared within the area of the lung in the prime, we reduced the possible target location distance from the centre from 240 pixels to 200. This randomly varied by a factor of 100 around this point and equally appeared either on the left or the right. As for Experiment 1a, to reduce anticipatory responses, we included further trials where no target appeared (catch trials) and each trial had a variable delay. Also, as for Experiment 1a, the images were flipped horizontally, so the right lung was positioned on the left side of the screen to mimic standard radiological presentation.

Procedure

After four practice trials per condition, the participants completed 192 experimental trials for each of the *subtle* condition followed by the *salient* condition (blocked; primes were randomly presented four times within each condition). A blank screen with fixation was displayed after the prime for each condition. In total, the participants saw 216 trials (192 experimental and 24 catch trials per condition) (See Fig. 5).

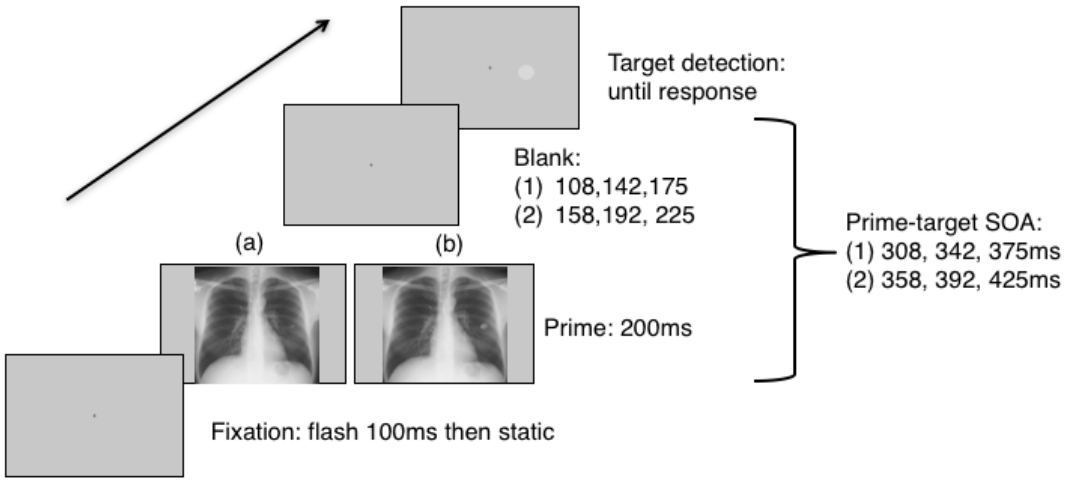


Figure 5: Example trial shown to 13 naïve participants in Experiment 1b. Trials began with a fixation cross. In separate blocks, the prime display was either (a), (b) corresponding to our *subtle* and *salient* nodule conditions. The prime was displayed for 200ms followed by a variable blank screen with fixation 108, 142 and 142ms or 158, 192 and 225ms (randomised within condition) followed by the simple dot detection (note: target larger and brighter for illustration purposes). The prime-target SOAs were 308, 342 and 375ms or 358, 392 and 425ms (blocked).

At the end of the experiment, the participants were asked to complete a 'nodule priors' questionnaire that consisted of a static image of a normal chest radiograph divided into quadrants (See Appendix B). First, they were asked to "Please mark 1-4 where you think the likely location for a single pulmonary nodule would occur, with 1 = most likely, 2 = likely, 3 = less likely, 4 = least likely". Second, we asked, "Do you know the frequencies of nodules in different areas?"

Results and Discussion

Analysis

Our dependent variable for target detection was reaction time (RT, milliseconds). Outliers (defined as RTs less than 100ms and greater than 1000ms) were removed prior to statistical analyses. With these criteria, 2% of trials were discarded across conditions. Catch-trial errors ranged from 2-11% across conditions and were removed from further analyses.

Subtle nodule condition. Prime-target SOA (1): Figure 5 (left two bars) shows the mean RTs (valid and invalid) for the *subtle* condition. A paired samples t-test confirmed that this difference was not significant [$t(12) = -.749$, $p = .47$, $M_{diff} = -4.95$, CI [-19.33, 9.44], $N = 13$, $r = .96$]. BF(13) = .35, providing strong support for the null hypothesis. *Prime-target SOA (2):* Figure 6 (middle, left 2 bars) shows the mean RTs for the *subtle* condition. A paired samples t-test confirmed that this difference was not significant [$t(12) = -.079$, $p = .94$, $M_{diff} = -.27$, CI [-7.42, 7.15], $N = 13$, $r = .97$]. As BF(13) = .28, we have support for the null hypothesis. These results provide evidence that naïve participants do not get cueing effects for the *subtle* nodules at prime-target onset delays of ~ 342 or 392ms.

Salient nodule condition. Prime-target SOA (1): Figure 5 (middle right two bars) shows the mean RTs (valid and invalid) for the *salient* condition. A paired samples t-test showed that this difference was not significant [$t(11) = -.526$, $p = .61$, $M_{diff} = 2.3$, CI [-7.33, 11.94], $N = 12$, $r = .98$]. BF(11) = .32 (Note: data from one observer were not recorded due to experimental error). *Prime-target SOA (2)*: Figure 6 (right 2 bars) shows the mean RTs for the *salient* condition. A paired samples t-test showed that this difference was significant [$t(12) = 2.53$, $p = .027$, $M_{diff} = 7.08$, CI [.98, 13.19], $N = 13$, $r =$

.99]. In standardised terms, this is a small effect (d_{unb} = .094, CI [.01, .19]). This shows evidence for a small, but reliable valid cueing effect for the *salient* nodules at a prime-target SOA of ~ 392, but not for ~ 342.

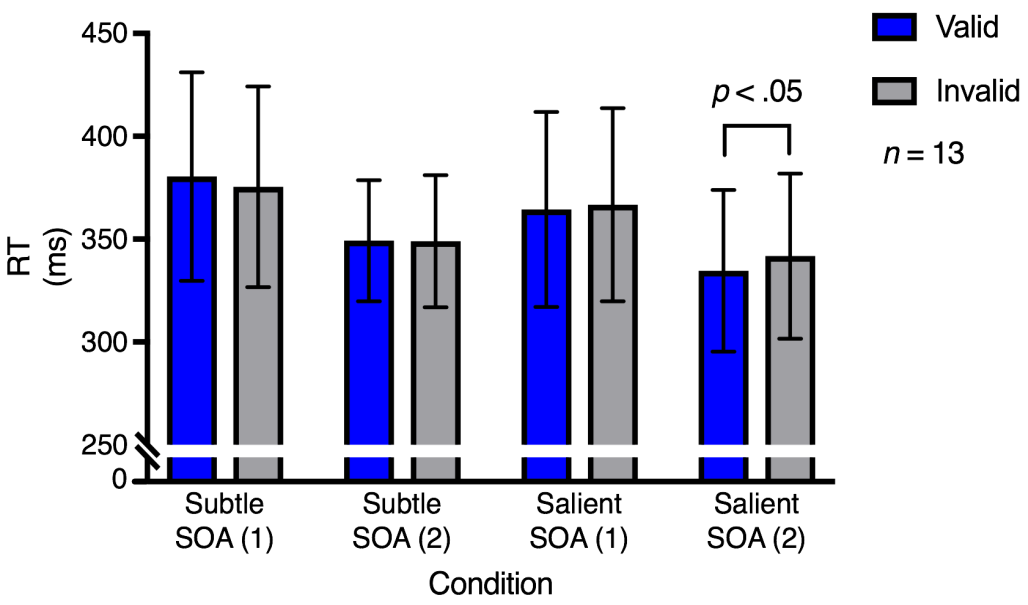


Figure 6: Experiment 1b. Mean Reaction Time (RT) for the naïve observers' performance on the cueing task for prime-target SOA of (1) 308, 342, 375ms, and (2) 358, 392, 425, for the *subtle* and *salient* nodule conditions. The blue bars represent the mean RTs for the valid trials and the grey bars represent the mean RTs for the invalid trials. Error bars represent 95% confidence intervals.

The results from Experiment 1, testing naïve observers, validate the experimental paradigm and identify the timings that can result in priming by showing a significant attention cueing effect for the *salient* nodules at a prime-target SOA of ~ 392ms.

Experiment 2 was designed to investigate a group of radiologists using similar timing parameters.

Results: Nodule Priors Questionnaire

We obtained responses from the naïve observers across these experiments, who have no previous experience reading medical images. The quadrant marked 'most likely' (assigned a '1') to contain a nodule was the upper left of the display (64%), followed by the upper right (22%), lower right (14%) and lower left (0%). None of the naïve observers reported explicitly knowing the frequencies of nodules in different areas of the lungs.

Experiment 2: Attentional cueing from medical images in radiologists

Our aim was to test whether invoking the context of a medical image would result in an attentional bias to the region of a chest radiograph most likely to contain a nodule (upper right quadrant). We also aimed to explore whether expertise boosted the salience of subtle signals (attentional capture by the original, subtle nodules that do not affect attentional allocation in naïve observers) and whether localisation information is extracted, as indexed by priming effects. Experts are thought to be sensitive to relevant features in an image and previous research suggests that brief durations are sufficient for experts to extract information about abnormalities (e.g., Carrigan, Wardle & Rich, (under review; Chapter 3); Evans et al., 2013; Kundel & Nodine, 1975). We therefore hypothesised that such information about the presence of nodule might drive a shift of attention, measurable by our priming paradigm.

Method

Participants

Thirteen radiologists (4 female, mean age = 50 years, SD = 10 years, range = 37 – 66 years; mean years qualified = 19 years, SD = 12 years, range = 2-36 years), from Sydney Metropolitan radiology practices volunteered and gave informed consent. All reported normal or corrected-to-normal vision and were naïve to the purposes of the experiment. The study was approved by the Macquarie University Human Research Ethics Committee (Medical Sciences).

Stimuli and Apparatus

For this experiment, we included only the *baseline, chest priors* and *subtle* prime conditions. As the salient primes already show an effect in naïve observers, including this condition was unnecessary. In addition, the radiologists' time was a valuable resource.

The experiment was conducted in in a quiet reading room in each radiologist's workplace, and was presented on a Macintosh MacBook Pro using MATLAB 2011B with Psychtoolbox Version 3 (Brainard, 1997; Pelli, 1997). Stimuli were downsized to 800 x 800 pixels centred on a 1920 x 1080 resolution 24-inch, LG W2442PA, liquid-crystal display screen, refresh rate of 60Hz. As for Experiment 1, the images were flipped horizontally, so the right lung was positioned on the left side of the screen to replicate the projection radiologists view in clinical practice. The participants sat at approximately 70cm away from the screen and the images subtended approximately 18° of visual angle and the target subtended approximately .7° of visual angle.

Procedure

There were three experimental conditions, blocked and presented in a set order. As the task was a simple target detection, to reduce the risk of set responses we included some jitter. For all conditions, the prime-target Mean SOA was 417ms (specific durations were 400, 417, and 450ms, randomised within condition). These timings were chosen after reviewing the RT data from Experiment 1b which indicated that the strongest cuing effects occurred at SOAs of 392-425ms. Apart from the timings, the baseline and chest priors conditions were identical to Experiment 1a and the radiologists performed 96 experimental trials, giving a total of 24 trials/condition (valid/invalid, left/right). The subtle nodule condition was identical to Experiment 1b (apart from the timings), and we increased the number of trials to maximise our sensitivity for detecting a subtle effect. The prime images were repeated so the radiologists saw the primes 4 times in each condition (valid = 2, invalid = 2), resulting in 192 experimental trials. To reduce anticipatory responses, we included a further 24 catch trials. For each of the conditions they saw 4 practice trials and breaks were given for 10 seconds every 48 trials. No feedback was provided (See Fig. 7). The radiologists were later sent the 'nodule priors' questionnaire (see Appendix B).

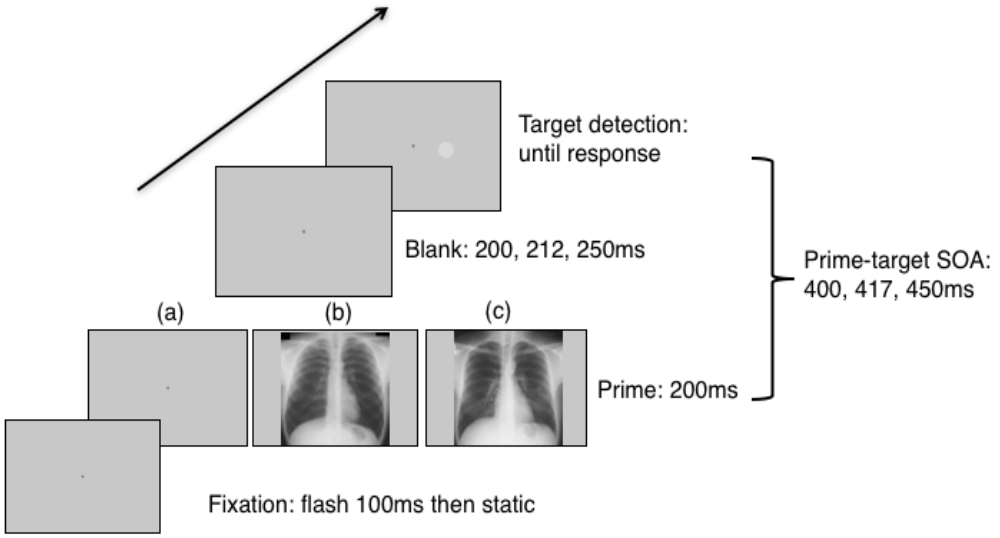


Figure 7: Example of an experimental trial shown to 13 radiologists in Experiment 2. Trials began with a fixation cross followed by the prime display. In separate blocks, the prime display was either (a), (b) or (c), corresponding to our *baseline*, *chest priors* or *subtle* nodule conditions. The prime was displayed for 200ms followed by a variable blank screen with fixation for 200, 212 or 250ms (randomised within condition) followed by the simple dot detection (note: target larger and brighter for illustration purposes). The target SOA was 400, 417 and 450ms.

Results and Discussion

Analysis

Our dependent variable for target detection was reaction time (RT, milliseconds). Outliers (defined as RTs less than 100ms and greater than 1000ms) were removed prior to statistical analyses. With these criteria 2-4% of trials were discarded across the three conditions. Catch-trial errors ranged from 1-2% and were removed from further analyses.

Baseline condition. When radiologists were asked to respond to the visual target with no prime, there was no difference in the RTs for left-sided versus right-sided targets. Figure 8a (left bars) shows the mean RTs for the baseline condition for left-sided target trials (blue bars) and right-sided trials (grey bars). A paired samples t-test showed this difference was not significant [$t(12) = -.501$, $p = .625$. $M_{diff} = 5.01$, CI [-16.78, 26.79], $d_{unb} = .04$, $N = 13$, $r = .95$]. As BF(13) = .31, there is support for the null hypothesis that the radiologists are equally fast at localising targets on either side. We interpret this as suggesting no underlying bias.

Chest priors condition. Lung nodules occur more frequently in the upper right quadrant (Winer-Muram, 2002). Figure 8a (right bars) shows the mean RTs when radiologists were asked to respond to the visual target following a normal chest radiograph prime for the left-sided target trials (blue bar) and right-sided target trials (grey bar). A paired samples t-test showed the difference in mean reaction time between the left and right cued trials was not significant [$t(12) = 2.107$, $p = .057$, $M_{diff} = 16.17$, CI [-.55, 32.88], $d_{unb} = .12$, $N = 13$, $r = .98$]. BF(13) = 1.46. This is another situation where a Bayes analysis can be very useful: our Bayes Factor of 1.46 shows that we have insufficient evidence to distinguish the null and alternative hypotheses; we would require more data to be able to draw conclusions.

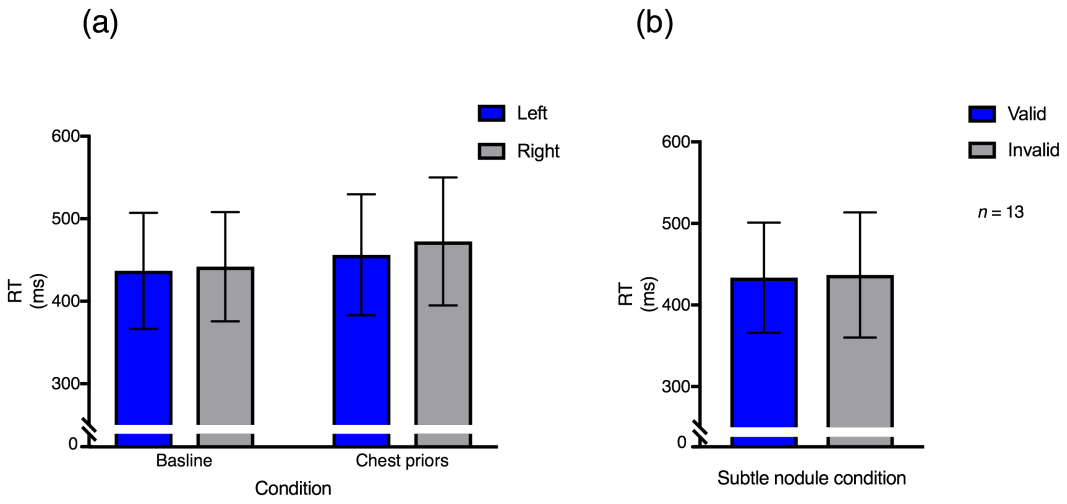


Figure 8: Experiment 2. Mean Reaction Time (RT) for the radiologists' performance on the cueing task. (a) Average RT for left and right trials for both the *baseline* and *chest priors* conditions; (b) Average RT for valid and invalid trials for the *subtle* nodule condition. Error bars represent 95% confidence intervals.

Subtle nodule condition: Figure 8b shows the mean RTs when participants were asked to respond to the visual target following a *subtle* prime for the valid trials (blue bar) and invalid trials (grey bar). A paired samples t-test showed that the participants were not significantly faster on valid vs. invalid trials [$t(12) = .66$, $p = .52$, $M_{diff} = 3.32$, CI [-7.64, 14.28], $d_{unb} = .03$, $N = 13$, $r = 1$]. BF(13) = .34. This provides support for the null hypothesis and indicates that, as a group, the radiologists do not show evidence for an attention cueing effect.

In many studies of expertise, the number of years of experience is an important factor. Here, we initially collapsed across all radiologists who fulfilled our criteria for

'experts'. In the sample, however, there is a lot of variability in years of experience, from 2-36 years. This lends itself to a post-hoc analysis to see if years of experience might correlate with any cueing effect. We note that any results of these post-hoc analyses need to be considered preliminary, particularly given the small sample size, and require replication before we can draw any firm conclusions. We conducted two correlation analyses to explore the relationship between experience and attention cueing.

First, we looked at the *chest priors condition* to see if years of experience affected the extent to which a normal chest radiograph influenced the deployment of attention. We calculated a hemifield effect in the direction of the hypothesis that attention will shift to the side of likely nodule location (that they will attend to the right lung (left screen) more than the left lung (right screen)) calculated as a difference score between left-sided target and right-sided target RTs. Figure 9a shows no relationship between years of experience and a left/right hemifield effect, Pearson's $r(13) = .033$, $p = .915$.

Next, we looked at the *subtle nodule condition.* To see whether there was a relationship between years of experience and cueing, we calculated a validity effect (calculated as a difference score between invalid and valid RTs). Figure 9b shows a significant positive correlation between years of experience and the validity effect, Pearson's $r(13) = .78$ $p = .002$. This suggests that the more experienced radiologists may be more sensitive to the subtle nodules than the less experienced radiologists.

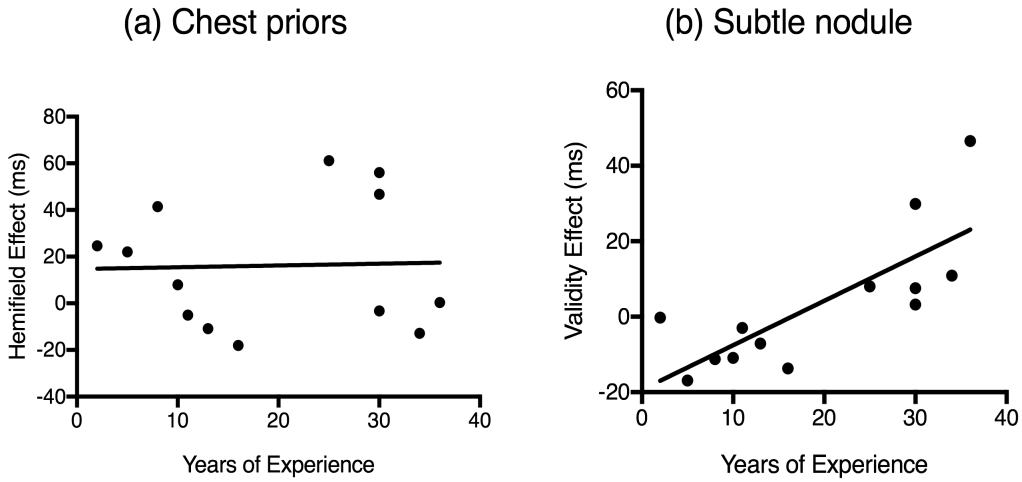


Figure 9: Correlation between years of experience (x-axis) for (a) the hemifield effect (y-axis) in the chest priors condition; and (b) the validity effect (y-axis) for the *subtle* nodule condition in Experiment 2.

Results: Nodule Priors Questionnaire

For the radiologists, 10/13 returned the questionnaire (retrospective to the experiment). The quadrant marked 'most likely' to contain a nodule was the upper right of the chest (50%), followed by the lower right (30%), upper left (20%) and lower left (0%). Somewhat surprisingly, none of the radiologists were aware of knowing the frequencies of nodules in different areas of the lungs. However, from this small sample we do see that half of the radiologists 'guessed' the upper right quadrant, which *is* the reported most frequent location of primary malignant nodules (Swensen, et al., 2000).

This suggests that there may be implicit knowledge of nodule locations, although a larger number of radiologist participants would be required to make any firm conclusions.

General Discussion

The successful completion of reading and interpreting medical images by radiologists is crucial for accurate diagnoses and patient care. These complex tasks reply upon the effective engagement of attentional mechanisms. Our first aim was to test whether normal chest radiographs presented as primes would result in a bias in radiologists' attention towards more likely locations for nodules. Second, for naïve observers we explored whether a *salient* cue in the context of a medical image caused a shift of attention. Finally, we investigated whether subtle cues are effective for radiologists due to their expertise with chest radiographs containing a nodule. We are using this as an indirect method for seeing if localisation information is present. In Experiment 1a, we tested naïve observers and as expected we did not show a bias in attention, but we also failed to show a cuing effect for the *salient* cues in a chest radiograph. In Experiment 1b, we adjusted the experimental timings and obtained a validity effect for naïve observers viewing a *salient* prime only, thus validating the paradigm. This was particularly important as exogenous cueing effects (such as are presumably occurring here in response to the salient nodule) typically occur around 100ms post cue and are short lived, lasting only for a few hundred milliseconds. In Experiment 2, for the radiologists, a Bayes analysis showed that we have inconclusive evidence that the 1.5-fold increase in nodules occurring the right vs. left lung (Swensen, et al., 2000; Winer, et al., 2002) influenced the initial distribution of attention in radiologists. We also failed to show a validity effect for a group of radiologists shown *subtle* primes, although a post-hoc analysis provides us with a hint that there may be a

correlation between experience and effects on our task. Given the replicated and robust effect of spatial attention cueing from a *salient* stimulus evident in classic laboratory paradigms (e.g., Posner, 1980), there are two possible explanations for the lack of the validity effect for the radiologists. First, is that the subtle nodules do not have a strong enough signal to capture attention in an exogenous sense. This could be because expertise does not tune feature sensitivity in that way, or, second, it could be there is an effect but our task in not sensitive enough to detect it.

We did not find conclusive experimental evidence either way on the question of whether the radiologists' prior knowledge about the location of nodules biased the allocation of attention. So, we cannot say for certain whether the information obtained is implied based on the statistical learning that accompanies experience. However, it is entirely possible that the variation in experience prevents us from seeing evidence for a bias. The evidence from the contextual cueing literature suggests that experience with the global context and spatial layouts of a display lead to a bias in attention deployment (Chun & Jiang, 1998). Further studies with a larger sample are required to investigate whether an experienced radiologist have developed an implicit memory for the likely location of a nodule in a medical image. This is critical because if a radiologist's allocation of attention is biased to a specific area, it implies that they might not attend to, and thus miss, an abnormality in a different region.

In Experiment 1b we showed evidence for an attention shift from a spatial cue (*salient* lung nodule) embedded in a cluttered medical image for a group of naïve observers. These findings are consistent with the strong bottom-up signal from the enhanced bright nodule among the lower contrast lung tissue capturing attention. Here, the naïve observers were unfamiliar with medical images and therefore would not be relying on context or prior knowledge to search for nodules. These results are useful as

they demonstrate that exogenous cueing can occur even when the cue is embedded within a cluttered heterogeneous prime image, which is quite different from most laboratory priming studies where factors such as distractors can be controlled (e.g., Chun & Jiang, 1998; Posner, 1980). Even though the outcomes for the radiologists (Experiment 2) are unclear, we have shown that cues embedded in medical images can drive the allocation of attention.

If further experiments show that highly experienced radiologists do indeed get a validity effect, this could reflect that they extract diagnostic information quicker in general from the radiograph, and therefore get the prime information (including nodule localisation) within a brief exposure. It could also be that the more, but not less experienced radiologists, are more attuned to the nodule and therefore it captures their attention to the location of the nodule in the prime. With many years of experience, this perceptual tuning we hypothesised might 'boost' the signal of the nodules in a top-down manner. Kundel & Nodine (1975) showed that when presented with a chest radiograph for 200ms experienced radiologists (but not trainees) could detect an abnormality with 70% accuracy. There are intriguing hints in our data that experience may be an important factor in the extent to which attention might be captured by suspicious-looking nodules. This would make sense for the perceptual tuning hypothesis of expertise where specific features in the image are attended to, and overall pattern recognition facilitates diagnosis (Nodine & Krupinski, 1998). Experienced radiologists have spent longer time reading chest cases and therefore have seen a higher number of images. To an expert, it is plausible that over time, a subtle signal might become more salient. We need further research to examine this hypothesis.

There are several limitations to the experiments presented in the chapter. First, we were constrained by the images so we not able to ensure that the nodule always appeared

within the same quadrant as the target. Thus, the prime and target were not always close in space (just on the same side). This means that even if the subtle prime shifted attention to its location, it might have been too distant from the subsequent target to facilitate detection. Our data from the naïve participants with salient versions of the same nodule locations, however, suggests that simply being cued to one side or the other with these displays can result in a priming effect. In future studies, it would be good to optimise the paradigm to get a larger effect in the naïve participants and then use these parameters for retesting a radiologist sample with subtle nodules. Another challenge when using medical images is the inherent variability that exists due to anatomical differences and image artefacts. At the outset, we chose to show the radiologists true pathology instead of the enhanced stimuli that the naïve observers saw, as we were concerned that the experts could 'spot a fake'. In addition, we wanted to maintain the ecological validity of the study by showing 'true' medical cases. Such variance in the prime images introduces another source of noise to the data, which could well mask any small effect. Third, the radiologists saw images slightly smaller in visual angle (and therefore possible decreased salience) due to the different presentation screen size compared with the naïve observers. Finally, the sample size in Experiment 2 was small. In practice, recruiting radiologists is challenging and obtaining larger sample sizes is difficult. Despite these limitations, this study is the beginning of an interesting new direction.

To our knowledge, this is the first study to investigate attention cueing of radiologists with medical images as stimuli, and it is as yet unknown if IOR and attentional cueing is moderated by expertise or experience when viewing medical images. A future study with a larger sample size, recruiting a broad range of experience including the sub-specialty of thoracic (chest) radiology, would be worthwhile as it would be interesting to see if there is a strong correlation between experience and cueing effects.

This would also allow for a more thorough investigation of the existence of chest priors (if indeed these exist) as our Bayes analysis was inconclusive, indicating we need more data to answer this question.

Overall, these experiments show promise. Here, we fail to see evidence for an attention shift in our expert group seeing *subtle* signals, but do when naïve observers see *salient* signals. We see intriguing clues that expertise could be related to the strength of cueing. Although further studies are required, this preliminary work suggests that within a medical imaging context, attention shifts *may* occur as a result of seeing a previously seen prime in highly experienced radiologists.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. doi:10.1163/156856897X00357

Biederman, I. (1972). Perceiving real-world scenes. *Science,* 177, 77–80.

Carrigan, A.J., Wardle, S.G., & Rich, A.N. (under review). Finding cancer in mammograms: if you know it's there do you know where it is? *Cognitive Research: Principles and Implications.*

Charness, A. N. (2014). The Role of Practice and Coaching in Entrepreneurial Skill Domains: An International Comparison of LifeSpanChess Skill. *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive psychology*, *36*(1), 28-71.

Evans, K.K., Georgian-Smith, D., Tambouret, R., Birdwell, R.L., & Wolfe, J.M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin Review, 20*(6)*,* 1170 – 1175. doi: 10.3758/s13423-013-0459-3

Evans, K. K., Haygood, T. M., Cooper, J., Culpan, A.-M., & Wolfe, J. M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences*, *113*(37), 10292–10297. http://doi.org/10.1073/pnas.1606187113.

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology-Human Perception and Performance*, *20*(2), 317-329.

Garland, L. H. (1961). Bronchial carcinoma—lobar distribution of lesions in 250 cases. *California medicine*, *94*, 7.

Holin, S.M., Dwork, R.E., Glaser, S., Rikli, A.E., & Stocklen, J.B. (1959). Solitary pulmonary nodules found in community-wide chest roentgenographic survey. *American Review Tuberculosis, 79*, 427-439.

Jiang, Y. V., Swallow, K. M., & Rosenbaum, G. M. (2013). Guidance of spatial attention by incidental learning and endogenous cuing. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 285.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. *Attention and performance IX*, *9*, 187-203.

Kundel, H.L., & Nodine, C.F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*(3), 527-532.

Kundel, H. L., Nodine, C. F., Krupinski, E. A., & Mello-Thoms, C. (2008). Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic radiology*, *15*(7), 881-886.

Lee, A. H. (2005). Why is carcinoma of the breast more frequent in the upper outer quadrant? A case series based on needle core biopsy diagnoses. *Breast, 14*, 151-2.

Midthun, D. E., Swensen, S. J., & Jett, J. R. (1993). Approach to the solitary pulmonary nodule. In *Mayo Clinic Proceedings*, *68*, 378-385.

Nodine, C. F., & Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Academic radiology*, *5*(9), 603-612.

Nodine, C.F., & Mello-Thoms, C. (2010). The role of expertise in radiologic image interpretation (p 139-156). In *The handbook of medical image perception and techniques*. Samei, & E. A. Krupinski (Eds.). Cambridge: Cambridge University Press.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3-25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 531-556). Hillsdale, NJ: Erlbaum.

Siegelman, S. S., Khouri, N. F., Leo, F. P., Fishman, E. K., Braverman, R. M., & Zerhouni, E. A. (1986). Solitary pulmonary nodules: CT assessment. *Radiology*, *160*, 307-312.

Shiraishi J, Katsuragawa S, Ikezoe J, et al. (2000). Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology, 174*, 71-74.

Swensen, S. J., Viggiano, R. W., Midthun, D. E., Müller, N. L., Sherrick, A., Yamashita, K., Naidich, D.P., Patz, E.F., Hartman, T.E., Muhm, J.R., & Weaver, A. L. (2000). Lung nodule enhancement at CT: multicenter study 1. *Radiology*, *214*, 73-80.

Winer-Muram, H. T., Jennings, S. G., Tarver, R. D., Aisen, A. M., Tann, M., Conces, D. J., & Meyer, C. A. (2002). Volumetric Growth Rate of Stage I Lung Cancer prior to Treatment: Serial CT Scanning 1. *Radiology*, *223*, 798-805.

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences, 15*(2), 77–84.

Chapter 5: General Discussion

General Discussion

The overarching goal of this thesis was to explore the cognitive processes that underpin important visual search tasks, such as those involved in medical image interpretation, focusing particularly on the early stages of vision. In a series of experiments, I explored the information that can be extracted from images regarding the presence and location of a target to reconcile two different theories of radiologist visual search (i.e., Evans, Georgian-Smith, Tambouret, Birdwell, & Wolfe, 2013; Kundel & Nodine, 1975). I used two stimulus types: natural scenes and naïve observers, and medical images on which I tested both novices and expert radiologists. Medical images and natural scenes share some key characteristics, such as inherent variability in the degree of clutter within a basic structure. In addition, it has been argued that we are all experts in perceiving scenes (Wolfe, Võ, Evans, & Greene, 2011), therefore using natural scenes and naïve observers forms a good model for radiologists with medical images. Thus, to address my key questions in this thesis I used both of these stimuli.

A major theoretical debate surrounds the extent to which radiologists have information about the location of a potential abnormality when they initially detect it. Although intuitively it seems that knowing there is *something abnormal* should also have information about *where that something is,* recent studies have claimed otherwise, suggesting a dissociation between detection and localisation of an abnormality (Evans, et al., 2013; Evans, Haygood, Cooper, Culpan, & Wolfe, 2016). These findings are inconsistent with a standard model of the stages of radiologist visual search (Kundel & Nodine, 1975). This thesis examines the claims by Evans and colleagues (2013; 2016) closely, and in this final chapter, I integrate the theoretical implications of my findings for the models of visual search.

I first review the specific findings from each experimental chapter and revisit the major issues which are addressed throughout my thesis. Then, I discuss some methodological considerations and factors that can affect study outcomes. Next, I discuss my study limitations and challenges. I then present the translational implications, which include clinical and broader implications. Finally, I discuss some outstanding questions and future directions.

5.1 Overview of findings

5.1.1 Chapter 2: The time course of rapid target detection and localisation

The experiments presented in Chapter 2, describe a behavioural paradigm where natural scenes were presented to naïve observers to investigate the amount of target detection and localisation information available after viewing a scene briefly. This design was an adaptation of a previously reported paradigm that presented medical images to expert observers (radiologists). I selected a Gabor as a target to avoid the complications of semantic relatedness to the scenes. Using scenes allowed an independent measure (scene categorisation) that 'gist' information was available at the selected durations. In two experiments, participants performed a two-alternate forced-choice task and were asked to detect and then localise a target presented within either Scene 1 or Scene 2 at presentation durations of 33-199ms (backward masked). In Experiment 1, the observers performed a precise localisation task (exact click) and a coarser task (left or right). I showed that a target could be detected and localised above chance for both tasks, even at the briefest duration, 33ms. In Experiment 2, using a similar procedure, I explored the influence of clutter during rapid presentations, operationalised as *enclosure* (Oliva & Torralba, 2001). Here, clutter affected performance in a predictable way: accuracy for both detection and localisation was higher for the scenes which were open compared with

closed, presumably due to the greater bottom-up salience of the target within the surrounding scene. The outcomes from this chapter extend what has been shown in the natural scene literature about object detection (e.g., Thorpe, Fize, & Marlot, 1996) and provide novel findings: localisation information is available after only seeing a natural scene for 33ms.

5.1.2 Chapter 3: Finding cancer in mammograms: if you know it's there, do you know where?

In Chapter 3, I adapted the paradigm used in the natural scenes experiments presented in Chapter 2 to explore the type of information available in medical images viewed by radiologists. Extending previous work by Evans et al. (2013), I first verified that a salient mass is able to be both detected and localised. These results confirmed that the experimental paradigm could measure these factors. I then did a second experiment replicating these findings with images more closely matched in difficulty to previous studies. In this study, I manipulated mammographic density, which increases the visual complexity of mammograms. With respect to the early visual processing of medical images, the work by Evans and colleagues (2013; 2016) has been predominant in the recent literature. These authors claimed that an initial global signal carries sufficient information to detect an abnormality, but this does not contain information about location. In my results, I found that radiologists could detect and localise a mass even for the most difficult condition (250ms, high density). I also showed that performance was mediated by density in a medical image in a predictable way: accuracy was lower for the high than low density images. I did, however, have trials on which detection was correct and localisation was incorrect, allowing me to fully investigate the possibility of a dissociation between detection and localisation. My approach emphasised the importance

of going beyond summary statistics to test this hypothesis, in particular to analyse the location errors at an image-level. Here, I discovered a number of factors that led to the underestimation of localisation including stimulus variability, response imprecision and participant guesses. Overall, my results from this experiment demonstrated that rapid localisation in mammograms is possible, and a normal patient variant, breast density (which contributes to the visual complexity of the image), has a large impact on radiologist search. I also outlined a method to fully test the account of Evans and colleagues (2013; 2016). I discuss this issue below in more detail.

5.1.3 Chapter 4: The influence of prior expectation and expertise on attentional cueing in medical images

In my final experimental chapter, I investigated whether priors about likely pathology locations guide attention, whether expertise boosts the salience of subtle nodules present in the lung tissue, and the extent to which location information is present, using a novel cueing paradigm where a chest radiograph (with or without a suspicious nodule) formed a prime. In two experiments, I presented prime chest radiographs to naïve observers and radiologists performing a simple visual detection task. After Experiment 1a showed effects that were influenced by masking, the modified design in Experiment 1b showed that an artificially-boosted nodule in the prime radiograph cued attention in naïve participants (valid < invalid reaction times), validating the task.

The results from Experiment 1b showed an effect of a salient but normal structure on naïve observers that did not seem to affect radiologists. Figure 1 shows an example of a normal anatomical variant commonly seen on a chest radiograph: a stomach bubble (red arrow). This darker area below the lungs is a gas bubble located in the fundus of the stomach. In naïve participants, there was a reverse validity effect which seems likely to

have been driven by a masking effect from the stomach bubble (right screen) when the target appeared on the same side. This effect did not occur for radiologists. This may be a hint that they are able to efficiently discount items that are salient in a bottom-up sense but are known to be irrelevant to the task based on expertise. Indeed, this is a hypothesis only that requires further exploration.

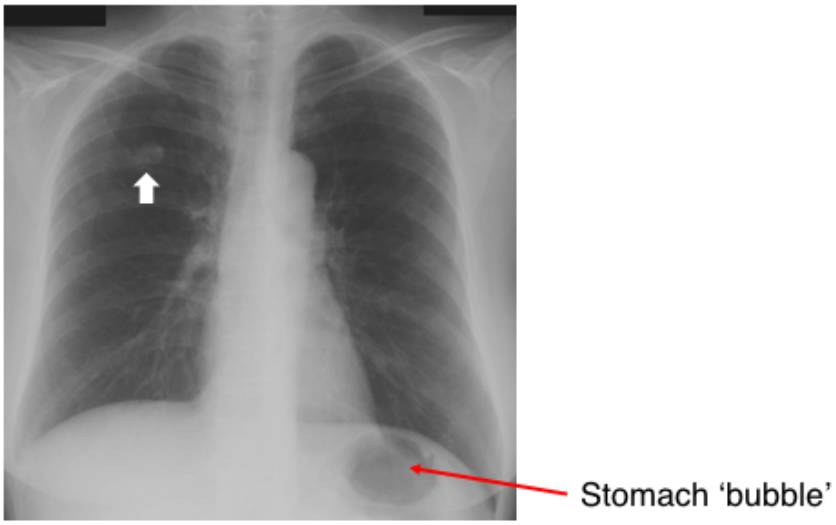


Figure 1: Example of a target-present chest radiograph presented to the participants in the subtle nodule condition from in Chapter 5. A subtle nodule (left screen, white arrow) is present in the lung tissue and the patient's stomach bubble is present on the lower right of the screen (red arrow).

In Experiment 2, with radiologists, we did not find conclusive experimental evidence either way on the question of whether attention shifted to locations based on statistical probability of a nodule. However, a post-hoc correlation analysis showed a significant relationship between experience and a validity effect: more experienced radiologists showed more effect of the prime than less experienced radiologists. This

suggests that expertise might boost sensitivity to the nodules. To my knowledge, this is the first study to consider the presence of priors and the effect of cueing in a medical imaging context, providing an indirect measure of the extent to which information about location is extracted by experts, and building upon the findings from the previous chapters. Future follow-up of this initial study is presented in a section 5.6.

The major issue that is addressed throughout my thesis is whether target localisation information can be extracted from brief displays, and whether it can be dissociated from detection. It seems intuitive that if you can detect a target you should also have information about its whereabouts. However, there are few studies which have investigated these two processes in the context of fast visual processing. Of these, some have argued that detection and localisation are dissociable and others have not found any evidence for localisation at all. As discussed in Chapters 2 and 3, using two different image sets, natural scenes and medical images, I found participants were able to both detect and localise targets above chance at comparable durations to previous work. In Chapter 3, I also found that apparent detection without localisation trials, particularly in difficult images, could be explained by other factors rather than a dissociation. When other possible explanations are taken into account (e.g., imprecision and guessing), there are few trials where detection is correct without localisation. These issues are discussed in more detail below. In Chapter 4, presenting chest radiographs as primes for 200ms, there were hints that perhaps highly experienced radiologists might get validity effects from subtle nodules, which would imply that localisation information must be present (although I note this was based on a post-hoc analysis and requires further study). Overall, my results do not provide evidence that detection and localisation are dissociable, suggesting instead that they are tightly coupled.

In any task where we present an observer with a stimulus display, there is an intricate interplay between top-down and bottom-up guidance of attention. Top-down guidance occurs when observers are looking for a specific target and can form an 'attentional set' for certain features. Bottom-up processing of the display will influence attention in terms of the way the stimulus interacts with the observer's visual system: high contrast or otherwise salient items in the display will tend to capture attention. In the medical images presented in Chapter 3, the radiologists 'set' their attention mechanisms to search for the features of an 'abnormal mass' present within a mammogram. Knowing the features of masses, a radiologist could use 'feature-based' guidance when searching an image. It is possible that the salience of a subtle nodule might be enhanced due to the effects of perceptual expertise. This suggests that there may be an interaction of top-down and bottom-up factors guiding attention in medical image search.

Many studies have shown that targets can be detected in the early stages of visual processing. In a classic rapid presentation study where observers performed a go/no-go categorization task, Thorpe et al. (1996) showed that observers could detect an animal/vehicle in a scene after seeing the display for 20ms. Although it is often cited that these objects can be detected in only 20ms, the lack of a backward mask makes the exact duration debatable. The Thorpe et al. (1996) study has potential issues which include the fact that the features of vehicles vs. animals are different. This means that observers could make the decision without actually seeing the whole animal, using a sense of whether there were more curves or straight edges in the display. My results from Chapter 2 show, with a stimulus that is completely arbitrary, that observers can indeed get enough information from brief displays to localise a target. These findings add to what is already known about objects in the natural scene literature about rapid target detection (e.g.,

Thorpe et al., 1996; VanRullen & Thorpe, 2001) by showing that a target can also be localised after a fast presentation.

My research has highlighted some important methodological issues that could affect the interpretation of prior results. Evans and colleagues (2013; 2016) briefly presented radiologists with mammograms (and patches from) and argue that abnormality *detection* but not *localisation* is based on the information provided by a global signal. There are a number of important issues with the Evans et al. (2013) and Evans et al. (2016) studies that my research addresses. I discuss each of these below.

5.2 Methodological considerations

5.2.1 Stimulus factors

For any experiment, the selected stimulus set is crucial and the use of real-world images comes with considerable challenges. First, the available images are often far from perfect for vision studies as it is difficult to control for factors such as co-existing variables (e.g., breast calcifications), target number and size, and breast tissue type. There is variability in medical images that make it possible that other factors could influence the results. In Chapter 2, I presented natural scenes as a model for expertise in visual search. Here, I could control for factors such as target number, size and location, enabling me to reduce the variability in the data. For my medical image experiments, I meticulously selected the stimuli to ensure that they were free of potential confounds, such as co-existing pathology and image artefacts. It is unclear whether this type of control has been taken in other studies (e.g., Evans et al., 2013; Evans et al., 2016).

The importance of considering the particular stimuli and responses at an individual image level was clearly illustrated in my response imprecision analysis in

Chapter 3. This analysis demonstrated coarse location information exists in many apparently 'incorrect' location responses. By looking at responses at the image level, I found that there were many instances where the radiologists' localisation responses were clustered around the correct area, thus providing evidence for coarse localisation information. This could be due to decay in the visual short-term memory trace over time or motor error in clicking the precise location. Individual images also showed instances where radiologists gave an apparently correct detection response but have then gone on to clearly localise a 'distractor' (see section 5.4 for further detail). Thus, selection of stimuli and analysis of the responses in detail is crucial to clear interpretation.

In clinical practice, there is a wide range of potential breast pathology which is required to be diagnosed. I deliberately selected images that contained a single mass, whereas Evans et al. (2013); Evans et al (2016) selected architectural distortions and subtle masses. This means that their stimuli may not have had a localisable signal or may have had multiple areas of abnormalities, making the localisation task unreliable and in some cases impossible. If there actually is not any artefact driving these effects (recall the artefact may only be present in a few images), then this is intriguing and evidence for a 'gist' or holistic signal of abnormality. However, this is not evidence for a dissociation between detection and localisation, because there may be nothing to localise on some images. We therefore need to ensure that future studies give full consideration to the potential impact of individual images on the results.

5.2.2 Summary statistics

Evans and colleagues' (2013; 2016) conclusions were made based on data that were very close to chance, where *d* prime was low. If the research question is, 'is detection (localisation) of an abnormality possible under these conditions?', as it probably

was when they commenced these studies, then looking for 'better than chance' performance makes sense. However, there are (at least) two issues that arise from the inference that detection greater than chance, and localisation not different from chance, equates to *no* localisation information. As this claim has significant theoretical implications, I consider these issues in detail below.

First, as discussed in Chapter 3, having *no* localisation is a null effect and therefore this cannot be interpreted using frequentist statistics. When a *p* value is greater than alpha this simply means that we have no evidence to reject the hull hypothesis. To claim that there is evidence for *no localisation*, appropriate statistics such as a Bayes Factor, would need to be calculated. This would then provide evidence of *no effect* (of localisation when there is detection) rather than just no evidence.

Second, summary statistics that are close to chance could be at risk of being influenced by a few images that might be driving the effect. If there are some images that are subject to any or all of the influences outlined in the above section on stimulus variables, this could drive a marginal effect. In Evans et al. (2016, Experiment 2), 14 radiologists were presented a single (thus far normal) mammogram for 500ms where the contralateral breast had a biopsy-proven cancer. They proposed that an abnormal signal may be present before the actual lesion is detectable by a radiologist, but do not elaborate on how this could be possible. Evans et al. (2016) are basing their findings on summary statistics where for the contralateral breast condition average $d' = .59$, which is only just above chance (d' chance = 0). For the ipsilateral side, average $d' = 1.16$, which was significantly higher than the contralateral condition. In a subsequent experiment, Evans et al. (2016) tested the hypothesis that if an abnormal signal is present across the entire breast then this signal should be present in isolated normal parenchyma. They presented radiologists 256- x 256-pixel lesion-free patches of a section from the abnormal breast

and lesion-free patches from the breast contralateral to the lesion. They propose that there is some signal in sections in both the ipsilateral and contralateral conditions because mean *d′* for detection was 0.33-0.4, which was significantly above chance. Given that the patches comprised of one eighth of the whole breast, they used a model of the whole breast which would show *d′* of 0.9-1.2. Both of these findings are intriguing as they claim that the radiologists were able to say something was abnormal based on a non-localised signal (as there was no abnormality to be localised) and means that these results could be influenced by a few images.

If one could be sure that image-level artefacts are not driving results in these studies, abnormality detection in contralateral and patches of non-lesion tissue may indeed provide evidence that 'gist' or other information at a global level can indeed support detection. They are not, however, evidence of a dissociation between detection and localisation, as there is effectively no localised mass or abnormality to detect in these experiments. They therefore do not provide any support for the basic claim that detection of a mass can be dissociated from location information, which is the crucial part of the argument, that gist can support detection.

5.2.3 Participant guessing

In any type of visual search study, on some occasions, the observers will make 'lucky guesses' which contribute to the data. Although these will be taken into account in a summary level *d′* analysis, any inference based on subsets of responses must consider this issue. To determine whether the trials where there was apparent 'detection without localisation' were indeed evidence of a dissociation, I tested whether a proportion of these could be explained by correct detection guesses. In Chapter 3, I calculated a

guessing probability based on a modified method described by Howe and Webb (2014). I created a hypothetical observer for each radiologist, one who only gets detection correct when it also has localisation information, meaning that any 'detection without localisation' trials are due to correct guessing. The prediction for each radiologist was compared with their observed performance and the results showed that there were actually very few trials representing a true dissociation between detection and localisation.

To summarise, my guessing correction results, the results of the response imprecision analysis, and the statistical factors discussed above, raise concerns about the interpretations presented in previous high-profile papers (Evans et al. 2013; Evans et al., 2016). My analyses identified the contribution of other factors to my (and potentially prior) results. It is crucial, particularly in research with such clear potential implications for clinical practice, that we, as scientists, are extremely sure of any claims we make.

Overall, it is certainly likely that in a normal visual search of medical images information perceived on the basis of gist contributes to successful diagnosis. However, as outlined above, there are important methodological issues to be considered before making the claim that such gist information can support detection in the absence of location information.

5.3 Models of visual search

The original model of the stages of radiologist visual search describes an initial stage where the extracted information (e.g., detection of an abnormality) then guides eye-movements in the second stage to the location of the abnormality (Kundel & Nodine, 1975). For this to occur, some information about location *must* be available in the initial

stage (Figure 2a). However, Evans and colleagues (2013; 2016) argue that the global signal *lacks* localisation information, rather providing only a sense that something is 'bad' in the image. On the basis of this sense, the search strategy of the radiologist then changes for a more thorough search for the source of that abnormality signal. Figure 2b shows the Evans et al. (2013) proposal, where a global signal changes search strategy.

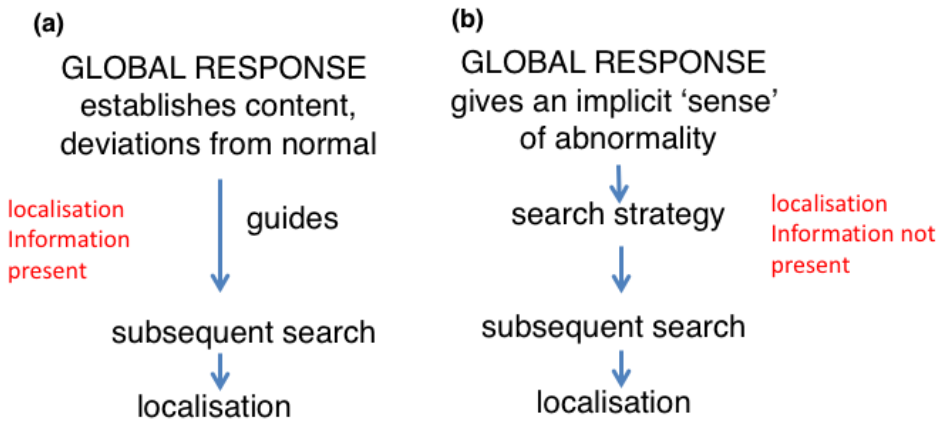


Figure 2. Basic representations of models of visual search in radiology. (a) Kundel and Nodine (1975); (b) Evans and colleagues (2013). The critical difference is whether the global response/signal contains localisation information (Kundel & Nodine, 1975) or not (Evans et al., 2013).

In this thesis, I replicate the effect that experienced radiologists can *detect* abnormalities in an image that is presented for 1 second or less, even within a quarter of a second. These findings are consistent with both of the above models and the medical perception literature (e.g., Kundel & La Follette, 1972; Kundel & Nodine, 1975; Kundel, Nodine, Krupinski, & Mello-Thoms, 2008; Nodine, Mello-Thoms, Kundel & Weinstein, 2002). The findings are also consistent with the two-pathway model proposed by Wolfe

et al. (2011) based on laboratory studies of basic visual search. In this model, the information picked up by the non-selective pathway includes structural cues but lacks the precision for object recognition. This 'gist' information then guides attentive search using a selective pathway that can achieve identification, presumably requiring accurate localisation.

Evans and colleagues (2013; 2016) seem to propose that in brief presentations of medical images, radiologists' detection could be achieved via the non-selective pathway. They base this claim on data showing detection occurring *without localisation*. They infer that correct detection must be based on gist – a global signal that could then change the search, rather than detection being based on any selective process. Indeed, this could explain the intriguing results where abnormalities in mammograms were detected in thus far normal parenchyma (i.e., no localisable signal) (Evans et al., 2016: Experiments 2 and 4). The findings presented in this thesis suggest that if an abnormality is detected, it can also be localised. This is consistent with the information supporting detection actually being within the selective pathway. This picture is consistent with the Kundel and Nodine (1975) model, which suggests information extracted in the first glance allows for detection and guides search to localising an abnormality. It is important to recall that these pathways are not strictly sequential, they feed into each other. Although the gist signal still exists through the non-selective pathway, I failed to find any evidence that this signal is what is driving successful detection.

The two-pathway model (Wolfe, et al., 2011) incorporates the proposal by Evans and colleagues (2013; 2016). Figure 3 shows the Wolfe et al. (2011) model with the conflicting findings illustrated. These results fit with the two-pathways working together, as they presumably do within the normal system, in the same way for medical images as for other visual searches. Thus, in the normal system, we would expect global (statistical)

information through the non-selective pathway whereas information in the selective pathway can support more detailed analysis including that required for object recognition and localisation.

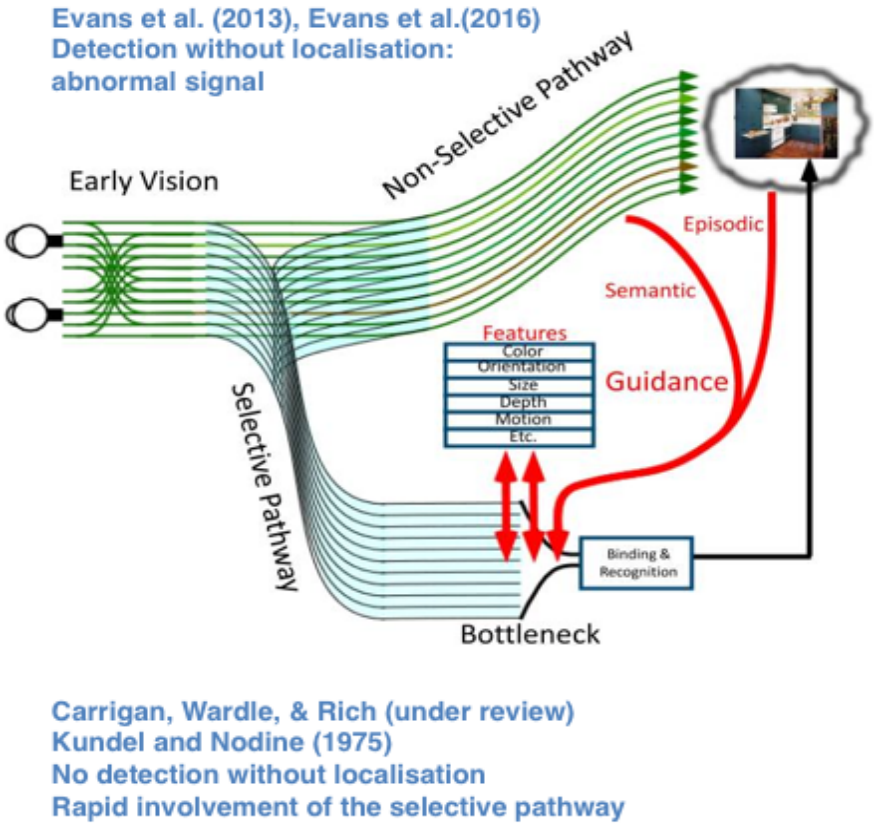


Figure 3. The two-pathway architecture for visual processing (from Wolfe, et al., 2011). Evans and colleagues (2013, 2016) propose that detection or an abnormal signal without localisation is achieved via the non-selective pathway. Carrigan et al. (under review) and Kundel and Nodine (1975) suggest that there is no detection without also having localisation information. This suggests rapid involvement of the selective pathway.

In Figure 4, I have built upon the theoretical model of Kundel and Nodine (1975) with modifications based on the evidence provided throughout this thesis. What others have described as a 'global response' is somewhat confusing because 'global response' and 'gist' have become synonymous in the literature. Therefore, it becomes unclear what is actually being discussed. Here, I have replaced this term with 'early processing' to reflect the findings in this thesis in a less theory-laden manner. After stimulus onset, at the level of the global response, there are three proposed mechanisms which work in parallel to lead to detection and localisation. Top-down goals such as knowledge of the target's features and the statistical likelihood of a target's location are postulated to influence the system at the same time as the bottom-up information such as salient features of the target. If the Evans et al. (2013) and Evan et al. (2016) proposal is subsequently found to hold, for example, under situations where the abnormality involves a more diffuse signal that could potentially modulate the gist, this would also contribute. Together this information guides subsequent search leading to localisation. This modified model aims to reconcile the previous experimental findings with those reported in this thesis.

When a stimulus has 'clutter' or higher complexity, there is more information to be processed through the system simultaneously. This results in less clearly defined bottom up information, reducing the benefit of this part of the guidance and therefore performance. It is well known that increased clutter or set size in a display degrades performance when observers are doing a free-viewing search task (Adamo, Cain & Mitroff, 2015; Asher, Tolhurst, Troscianko, & Gilchrist, 2013; Rosenholtz, Li, Mansfiled & Jin, 2005; Rosenholtz, Li, & Nakano, 2007; Whitney & Levi, 2011; Wolfe, 1994). In Chapters 2 (Gabor in natural scenes) and 3 (mass in mammograms), I showed that when the scene or mammogram is complex (closed scenes or high density mammograms),

performance is reduced relative to the simpler displays (open scenes or low density mammograms), likely because the target is less salient among the cluttered background. Presumably in this situation, the observer has to rely more on the other mechanisms in play.

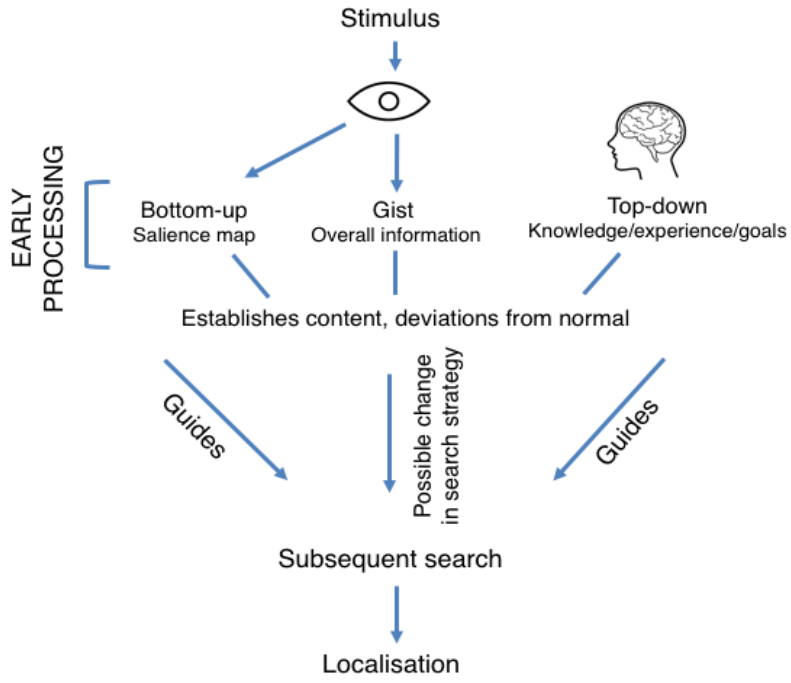


Figure 4. Modification of Kundel and Nodine's (1975) model of the stages of radiologist visual search. At the level of early processing, attentional deployment presumably reflects the input from both bottom up and top down factors. The type of 'gist' analysis described by Evans et al. (2013) (included in the non-selective pathway by Wolfe et al. (2011)) would contribute to perception, but the overall outcome of the global response in early processing is to identify deviations from normal that can guide subsequent search to the abnormality location.

5.4 Limitations & challenges

As discussed in section 5.2.1, the real-world images have significant challenges. The limitation of meticulously selecting the stimuli means that to obtain an adequate number of trials within an experiment, in some cases the images were repeated (e.g. chest radiographs in Chapter 4). The mammograms used in Chapter 3 were selected to exclude those images which contained potential confounding variables. Figure 5 illustrates examples of images excluded from the stimuli set. Figure 5a is an example of an image that was excluded due to a large, centrally located breast calcification, which is a normal variant often seen among the breast tissue (bright white area, red arrow). Figure 5b is an example of an image which contained a small calcification (bright white area, red arrow) along with a normal axillary lymph node (red arrow). These naturally occurring variants would distract the radiologists, potentially leading to misleading detection responses (cf. Chapter 3, Fig.8). However, the need for such strict selection resulted in limited image sets.

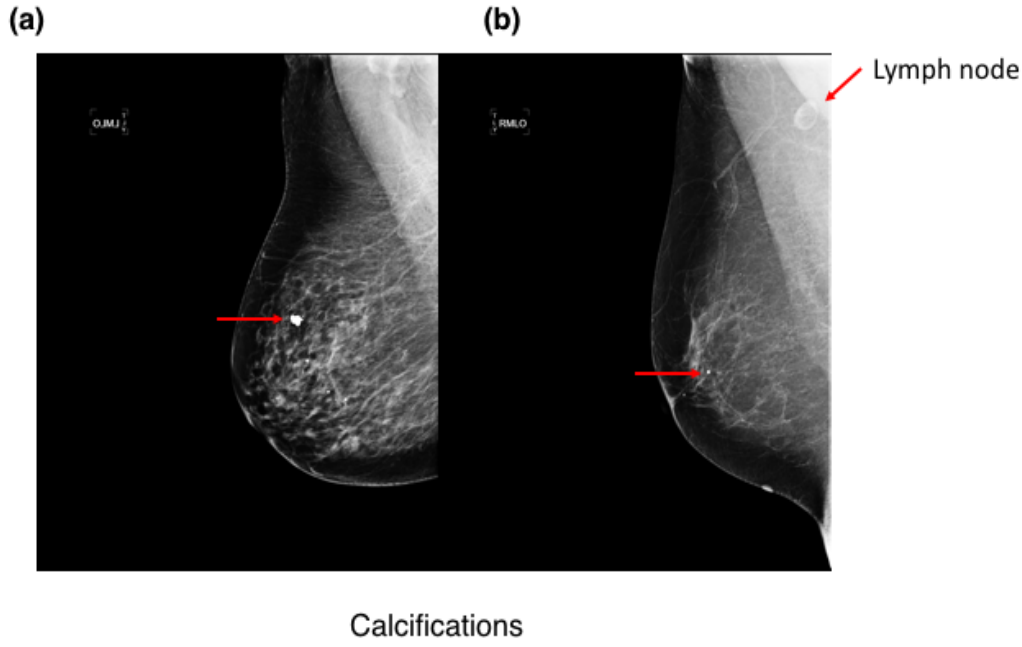


Figure 5: Examples of two mammograms which were excluded from the stimulus set from the experiments presented in Chapters 3 and 4 as they contained distractor items. (a) A large, centrally located breast calcification (red arrow) which appears salient in the image; (b) A smaller calcification (red arrow) and also a normal axillary lymph node.

Despite applying a strict selection criterion, the post-hoc inspection of the images revealed that there were still characteristics in a few images that were naturally occurring stimulus features, which may affect the results. In Chapter 3, I delved into response errors by looking closely at the responses to each image. This image-level analysis made it clear that in some apparent correct 'detection without localisation' trials, radiologists were actually detecting a distractor as a target and these trials contributed to the overall summary data of 'detection without localisation'. As discussed above, this is a clear concern for interpretation of summary statistics alone, and suggests image-level analysis is crucial for studies wishing to access potential dissociations.

In Chapter 4, chest radiographs were selected as these stimuli are more symmetrical than mammograms, which was crucial to test for cueing in for this paradigm. A set of chest radiographs downloaded from the Japanese Society of Radiological Technology database (JSRT: Shiraishi, Katsuragawa, Ikezoe, Matsumoto, Kobayashi et al., 2000) were selected as the best options, excluding those with image artefacts. As the radiographs were digitised images from analogue film this meant many contained artefacts such as anatomical variants or marker opacities and even tape and pencil markings which could act as distractors. It is possible that these stimulus factors may have cluttered the image with features that are more salient than the actual nodule, reducing the potential for capturing attention to the nodule.

Excluding images also meant a reduction in the number of stimuli in the sets for all the experiments presented in Chapters 3 and 4. To overcome this, in Chapter 3, Experiment 2, I tested only two duration blocks (instead of the 3 presented in Chapter 3, Experiment 1) to maximise the number of trials per condition, as I had introduced a further factor (density). In Chapter 4, I increased the number of primes by randomly repeating each one the same number of times to also increase the number of trials per condition. As the subsequent target location was randomised, the prime/target was not repeated in exactly the same location, so contextual cueing is unlikely. It is also unlikely the observers remembered the prime after a repeated exposure, given the brevity of the displays. However, overall, it would be ideal to have many more unique images per experiment to avoid any such potential influences.

A further limitation when using medical images as stimuli is that I was unable to control for nodule location within each quadrant and could only balance nodule number across the vertical axis. This means that the prime was not always close in space to the target, decreasing the likelihood I would see evidence of any attention shift that did occur.

A future study with tighter control on prime-target locations may be more successful in detecting cueing effects, which are likely to be quite subtle.

A challenge when studying experts is that it is often difficult to obtain a large enough sample size for adequate statistical power. Here, in addition to recruiting radiologists for my key studies, I also used a model of expert visual search using natural scenes. Based on the scene perception literature (e.g., Oliva & Torralba, 2001), I was able to obtain exemplars easily and categorise these at the superordinate and basic-level. Although Chapter 2 is not about scene perception *per se*, I selected these as stimuli because they form a category with which we all have extensive perceptual experience. In addition, they provide a way of independently verifying that 'gist' is extracted. The main advantage of using scenes was that I was able to maintain experimental control of target location within the scene (unlike Evans and colleagues (2013; 2016) and my medical image studies). Using a target such as a Gabor also allowed me to control factors such as size, location and semantics, so the observers would not be relying on scene guidance to make judgements (e.g., Wolfe, et al., 2011). By presenting scenes, I was also able to recruit an adequate number of observers from the Macquarie University participant pool to maximise power.

Sample size was a limitation for the attentional cueing study discussed in Chapter 4, Experiment 2. Here, I was only able to recruit thirteen radiologists for the study, and they had a wide range of experience (2 – 30+ years). This may have been an issue as level of experience could be a key factor in determining potential cueing. The range did, however, give us the opportunity for an exploratory post-hoc analysis assessing the relationship between the magnitude of the validity effect and the years of experience. This analysis showed a significant positive correlation. For the less experienced radiologists, it is entirely possible that their perceptual expertise, and thus recognition of

subtle nodules from such a brief (200ms) image, had not yet developed. This direction is promising and this experiment provides important pilot data for a future larger study.

5.5 Translational implications

5.5.1 Clinical implications

My research has implications for clinical practice in Australia. As outlined in Chapter 3, mammographic breast density (MBD: Li, et al., 2013) is an important factor to consider in a medical imaging context. My research findings converge with clinical research outcomes of increased risk and decreased diagnostic sensitivity to suggest MBD is a major factor in diagnostic performance. MDB varies widely and approximately half of the female population have dense breasts (Sprague, et al., 2014). As there is more functional breast tissue present in a dense breast, there is an increased risk of developing cancer in such breasts (Boyd, et al., 2010). Studies have also shown that high levels of breast density reduce radiologist sensitivity in free-viewing (see Al-Mousa, Ryan, Mello-Thoms & Brennan, 2014). In Chapter 3, my results confirm that the normal anatomical variant of breast density seen in patients affects early visual search processes. Whereas previous studies only demonstrated that high levels of breast density impede radiologist diagnostic performance, my research demonstrates that breast density has a negative effect on the information a radiologist is able to extract from the first processing stages (the ‘first glance’). Given it has been proposed that diagnostic decisions depend crucially on the information available during this time (Kundel et al., 2008; Kundel & Nodine, 1975; Evans et al., 2013; Evans et al., 2016), these results provide evidence that density is an important consideration for clinical practice.

Breast density is a factor which is not currently recorded in patient screening reports in Australia. If referring clinicians and patients were notified of having breasts which contain high levels of MBD then this would lead to the provision of more suitable imaging modalities such as 3D mammography (digital breast tomosynthesis: DBT), ultrasound or magnetic resonance imaging, and thus an earlier cancer diagnosis. It is well known that the key to cancer survival is early diagnosis. Along with teaching trainee radiologists, notification for both patient and clinician about breast density and potential cancer risk may have a significant positive effect on outcomes.

With the advances in computerisation and technology, medical imaging is undergoing rapid change. The development of digital breast tomosynthesis (DBT), which provides volumetric 3-dimensional image data in mammography, aims to reduce tissue overlap that often accompanies a mammogram. DBT is a reconstructed mammogram taken from many 'slices' through the breast parenchyma (Poplack, Tosteson, Kogel, & Nagy, 2007) and is often used in a screening context for patients with complex breast tissue such as dense breasts (Alakhras, Brennan, Rickard, Bourne, & Mello-Thoms, 2015). Research has shown that the addition of DBT potentially increases cancer detection, and reduces false positives and recall rates (Alakhras et al., 2015; Poplack et al., 2007). If density is not routinely reported in screening context, then the referring clinician and patient are not made aware that further imaging such as DBT may be warranted.

There is a global endeavour to understand the complex issues associated with applied visual search, how the processes work, and what strategies are beneficial for boosting performance. In radiology, a better understanding of the cognitive mechanisms underpinning visual search will lead to improved diagnostic performance, and better health outcomes. My research has implications for policy and radiologist training, in the

need for breast density to be mandatorily reported in mammographic screening, and for greater consideration in training.

5.5.2 Broader implications

The research presented in this thesis falls squarely into Pasteur's quadrant of use-inspired basic research (Stokes, 2011). In this thesis, I have taken a real-world problem of cancer diagnosis and applied fundamental cognitive science tools to understand the basis of initial processing of medical images by radiologists. For example, by using natural scenes as a model for medical image search, I have linked two literatures and my results can therefore inform basic vision science as well as the medical community. Most papers in this field stop at the level of summary statistics (e.g., average accuracy or reaction time), whereas in other fields, like language-research, an item-based analysis is common. My research emphasises the importance of the image-based analysis (e.g., Chapter 3,) which allowed important insights, and indeed, counter-evidence to a high-profile claim about the mechanism underpinning radiologist search. This thesis also builds upon the literature surrounding early visual processing and expertise. I investigated factors relating to visual search and attention and related these to the real-world context of medical imaging. These results contribute to a growing literature where rigorous science techniques have been applied successfully to address important clinical questions.

The findings presented in this thesis also have implications for several other similar real-world situations where experts need to examine complex images. In the areas of airport security and military surveillance, for example, understanding how people are able to detect and localise information efficiently is crucial for public safety.

This research could inform computer-aided classification of abnormalities. A conventional classification system used widely in clinical practice is computer-aided detection (CAD). CAD is a tool which aims to increase the detection rate and thus reduce errors in digital radiology (Al Mohammad, Brennan & Mello-Thoms, 2017). It requires several steps which include pattern recognition and image processing (Hua, Hsu, Hidayati, Cheng, & Chen, 2015). The algorithms developed for CAD are based on the features presumed to be use when a radiologist is searching an image, namely pattern recognition (Castellino, 2005). An expert radiologist has learned to visually recognise normal vs. abnormal patterns and can do so rapidly (Drew, Evans, Võ, Jacobson & Wolfe, 2013; Kundel & Nodine, 1975; Nodine & Krupinski, 1998). In a machine learning context, CAD is also focusing on pattern recognition and regularities/irregularities in the data. The findings in this thesis, specifically the features detected and localised in the first glance, could inform CAD. They have contributed to the understanding of the information radiologists are basing their decisions on, and could inform trainees who are in the process of developing visual expertise.

The development of diagnostic tools raises important questions: what impact will recent computational advances such as deep neural networks have on medical imaging? One of the main criticisms of CAD is the large amount of processing required to differentiate malignant and benign features. To reach a 'decision' the steps involved include feature computing, selection and integration - every subsequent step rests heavily on the success for the prior stage. Recently, machine-learning techniques, such as deep learning, have been explored in the context of tumour classification in radiology (Hua, et al., 2015). In deep learning, using a hierarchical structure, a number of layers of data are exploited for pattern classification and feature learning. Hua et al. (2015) showed that these methods achieved better pulmonary nodule classification than CAD in computer

tomography (CT). Overall, these systems show promise, reducing errors and false alarms, and improving diagnostic outcomes. A greater understanding of human diagnosis will provide important foundations for future development of automated tools. We also need further research studying the interaction between humans and automated diagnostic tools.

5.6 Future directions

The research in this thesis opens up several clear research directions. For the experiment presented in Chapter 3, a future study with a larger sample size, presenting pathologically varied mammograms (i.e. not only single masses) would be important to ensure results replicate and generalise to larger samples. The preliminary results presented in Chapter 4 offer intriguing hints that priming may indeed be possible by subtle lesions for highly experienced radiologists. This is likely to be a small effect, however, and requires a large sample to fully test the idea. This also offers another method for looking at the localisation question, and may contribute to the notion of developing expertise. Further exploration of the influence of prior knowledge will also be important as it is possible that these priors affect the allocation of attention. A follow-up experiment should also source a higher number of chest radiographs that contain nodules balanced across the four quadrants of the lung to constrain prime-target locations more closely. These experimental directions follow from the research presented in this thesis.

Radiology is moving from 2D towards more 3D modalities (e.g. DBT, CT), and we know very little about visual search in a 3D context within medical imaging. Currently, cognitive psychology paradigms are in development which can be used for both non-expert and expert populations. For example, a segemented-3D search paradigm that matches the critical qualities of digital breast tomosynthesis allows presentation of

stimuli that appear ‘mammogram like’, but using target ‘Ts’ among ‘Ls’ arranged throughout (Adamo, personal communication, August, 2017). What is unique about this paradigm is the ability to create a large number of exemplars and the ease with which control and flexibility to manipulate factors such as target number, and set size in the search display is achieved. This overcomes some of the challenges using real-world images and could provide useful insights into expert visual search, especially pertaining to rapid target detection and localisation.

It is also crucial that we test radiologists in conditions more closely aligned to radiology practice. Somewhat related to priming effects are the errors caused by the phenomena known as the *attentional blink* (AB: Broadbent & Broadbent, 1987; Raymond, Shapiro, & Arnell, 1992). The AB is a temporal search phenomenon where the second of two targets is missed when it appears close in time to the first. AB is studied during a rapid-serial-visual-presentation (RSVP) paradigm where items are sequentially displayed briefly and observers are required to detect targets. In medical imaging an enormous amount of data is viewed within a 3D reconstructed examination. For instance, in a typical abdominal CT scan ~1000 images are generated (Reiner, Siegel, & Siddiqui, 2003). To view a scan the radiologist selects a case and then scrolls through the images with a mouse. Each radiologist is likely to adopt a scrolling speed/method which is behaviourally suited for their diagnostic purpose. One study reported that the average scrolling rate through CT images for radiologists is around 25-30 frames/second, or 40ms/frame (Diaz, Schmidt, Verdun, & Bochud, 2015). Depending on the individual’s speed and the location/size of targets, it is plausible that targets may be missed, especially when a second one appears temporally close to the first. Unanswered questions include: does the AB similarly occur with 3D volumes? Does the magnitude of the AB correlate with levels of experience of a radiologist? It is important that we investigate the extent to

which the AB occurs when radiologists view rapidly presented displays scrolling through a 3D search space.

The recent, rapid and continual growth in computational and 3-dimensonal modalities is providing fertile ground for future studies. In a broader context, there remain many unanswered questions in the medical perception research literature. These include how can we, as vision scientists, inform sophisticated machine learning techniques about the cognitive underpinnings of the processes involved in the human search engine? This also poses the contentious question: will the human search engine be replaced in the future?

5.7 Conclusions and final remarks

The findings in this thesis advance our knowledge about early visual processing and, consistent with previous research, show that a remarkable amount of information can be processed from very brief displays. Specifically, when observers view a natural scene or medical image briefly, they can extract not just information allowing detection of a target, but this is tightly linked with information about target location. I have also examined important factors in medical image perception, including how the normal anatomical variant of breast density affects early visual search.

My research approaches a real-world problem of detection of abnormalities in medical images using rigorous methods of cognitive science. By using natural scenes as a model for medical image search, I have linked two literatures and my results can therefore inform basic vision science as well as the medical community. I have applied fundamental cognitive science tools to understand the basis of initial processing of medical images by radiologists. My approach emphasised the importance of going

beyond the summary statistics when interpreting findings. For instance, a thorough analysis at the image level has important implications for the way researchers are studying questions of localisation in radiology. By taking this approach, this research advances the scientific understanding of the perceptual limitations in medical imaging.

Overall, my work presented in this thesis is important in the scientific arena by promoting good practice and addressing important theoretical questions about how we search complex images. The implications from these findings are broad: for the medical and health arena, with direct policy and training implications, and for the general public.

5.8 References

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2015). Targets Need Their Own Personal Space: Effects of Clutter on Multiple-Target Search Accuracy. *Perception*, *44*(10), 1203–1214. http://doi.org/10.1177/0301006615594921.

Alakhras, M. M., Brennan, P. C., Rickard, M., Bourne, R., & Mello-Thoms, C. (2015). Effect of radiologists' experience on breast cancer detection and localization using digital breast tomosynthesis. *European radiology*, *25*(2), 402-409.

Al Mohammad, B., Brennan, P. C., & Mello-Thoms, C. (2017). A review of lung cancer screening and the role of computer-aided detection. *Clinical Radiology*, *72*(6), 433-442.

Al Mousa, D. A., Ryan, E. A., Mello-Thoms, C., & Brennan, P. C. (2014). What effect does mammographic breast density have on lesion detection in digital mammography? *Clinical radiology*, *69*(4), 333-341.

Asher, M.F., Tolhurst, D.J., Troscianko, T., & Gilchrist, I.D. (2013). Regional effects of clutter on human target detection performance. *Journal of Vision*, *13*(5)*:25*, 1-15.

Boyd, N. F, Martin, L. J, Bronskill, M, Yaffe, M. J, Duric, N & Minkin S. (2010). Breast tissue composition and susceptibility to breast cancer. *J Natl Cancer Inst, 102*(16), 1224-1237.

Broadbent, D., & Broadbent, M. (1987). From detection to iden- tification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*, 105–113.

Castellino, R. A. (2005). Computer aided detection (CAD): an overview. *Cancer Imaging*, *5*(1), 17–19. http://doi.org/10.1102/1470-7330.2005.0018

Diaz, I., Schmidt, S., Verdun, F. R., & Bochud, F. O. (2015). Eye-tracking of nodule detection in lung CT volumetric data. *Medical physics*, *42*(6), 2925-2932.

Drew, T., Evans, K., Võ, M. L. H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics*, *33*(1), 263-274.

Evans, K. K., Georgian-Smith, D., Tambouret, R., Birdwell, R.L., & Wolfe, J.M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin Review.* doi: 10.3758/s13423-013-0459-3

Evans, K. K., Haygood, T. M., Cooper, J., Culpan, A.-M., & Wolfe, J. M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences*, *113*(37), 10292–10297. http://doi.org/10.1073/pnas.1606187113.

Howe, P.D.L. & Webb, M.E. (2014). Detecting unidentified changes. *PloS ONE, 9*(1), e84490.

Hua, K. L., Hsu, C. H., Hidayati, S. C., Cheng, W. H., & Chen, Y. J. (2015). Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and therapy*, *8*.

Kundel, H. L., & La Follette Jr, P. S. (1972). Visual search patterns and experience with radiological images. *Radiology*, *103*(3), 523-528.

Kundel, H. L., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology*, *116*(3), 527-532.

Kundel, H. L., Nodine, C. F., Krupinski, E. A., & Mello-Thoms, C. (2008). Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic radiology*, *15*(7), 881-886.

Li, J., Humphreys, K., Eriksson, L., Edgnen, G., Czene, K., & Hall, P. (2013). Mammographic density reduction is a prognostic marker of response to adjuvant tamoxifen therapy in postmenopausal patients with breast cancer. *Journal of Clinical Oncology, 31*(18), 2249-2256.

Nodine, C. F., & Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Academic radiology*, *5*(9), 603-612.

Nodine, C. F., Mello-Thoms, C., Kundel, H. L., & Weinstein, S. P. (2002). Time course of perception and decision making during mammographic interpretation. *American Journal of Roentgenology*, *179*(4), 917-923.

Oliva, A. & Torralba, A. (2001). Publication utilizing eight semantically organized categories: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. From http://cvcl.mit.edu/database.htm

Poplack, S. P., Tosteson, T. D., Kogel, C. A., & Nagy, H. M. (2007). Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. *American Journal of Roentgenology*, *189*(3), 616-623.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 849–860.

Reiner, B. I., Siegel, E. L., & Siddiqui, K. (2003). Evolution of the digital revolution: a radiologist perspective. *Journal of Digital Imaging*, *16*(4), 324-330.

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005, April). Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 761-770). ACM.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17, 1-22. http://doi.org/10.1167/7.2.17.

Shiraishi J, Katsuragawa S, Ikezoe J, et al. (2000). Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology, 174*, 71-74.

Sprague, B. L., Gangnon, R. E., Burt, V., Trentham-Dietz, A., Hampton, J. M., Wellman, R. D., Kerlikowske, K., & Miglioretti, D. L. (2014). Prevalence of mammographically dense breasts in the United States. *Journal of the National Cancer Institute*, *106*(10), dju255

Stokes, D. E. (2011). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582*),* 520 – 522.

VanRullen, R., & Thorpe, S.J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artefactual objects. *Perception, 30*(6), 655-688.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, *15*(4), 160-168.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.

Wolfe, J.M., Võ, M.L., Evans, K.K., & Greene, M.R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Science, 15*(2), 77-84.

Appendix A: Finding cancer in mammograms: if you know it's there, do you know where?

Ann. J. Carrigan[1,2,3], Susan. G. Wardle[1,2], Anina. N. Rich[1,2,3].

[1]*Perception in Action Research Centre & Department of Cognitive Science, Macquarie University, Australia*

[2]*ARC Centre of Excellence in Cognition & Its Disorders, Macquarie University, Australia*

[3]*Centre for Elite Performance, Expertise, and Training, Macquarie University, Australia*

First and corresponding author: ann.carrigan@mq.edu.au
Second author: susan.wardle@mq.edu.au
Last author: anina.rich@mq.edu.au

The manuscript that describes this study was co-authored by Carrigan, A.J., Wardle, S.G., & Rich, A.N. (under revision following review). Finding cancer in mammograms: if you know it's there, do you know where it is? *Cognitive Research: Principles and Implications*.

Abstract

Humans can extract considerable information from scenes, even when these are presented extremely quickly. The ability of an experienced radiologist to rapidly detect an abnormality on a mammogram may build upon this general capacity. Although radiologists have been shown to be able to detect an abnormality 'above chance' at short durations, the extent to which abnormalities can be localised at brief presentations is less clear. Extending previous work, we presented radiologists with unilateral mammograms, 50% containing a mass, for 250 or 1000ms. As the female breast varies with respect to the level of normal fibroglandular tissue, the images were categorised into high and low density (50% of each), resulting in difficult and easy searches respectively. Participants were asked to decide whether there was an abnormality (detection) and then to locate the mass on a blank outline of the mammogram (localisation). A post-hoc analysis showed both detection and localisation information for all conditions when we include a wider acceptance localisation boundary. Although there may be a dissociation between detection and localisation on a small proportion of trials, we find a number of factors that lead to the underestimation of localisation including stimulus variability, response imprecision and participant guesses. We emphasise the importance of taking these factors into account when interpreting results. The effect of density on detection and localisation highlights the importance of considering breast density in medical screening.

Keywords: visual search, medical imaging, global processing, breast density, target detection, target localisation

Significance

In medical imaging, a radiologist searches and interprets a medical image to make critical diagnostic decisions (e.g., is that a cancer or not?), often under time pressure. With time and practice, experienced radiologists are thought to develop skills that allow them to form the basis of a diagnosis (normal or abnormal) during an initial glance at an image. This implies that the information extracted from the image in the first second of processing contains critical information that informs diagnosis. Here, we explore what type of information is present in this timeframe, particularly focusing on the presence (or lack thereof) of information about the location of potential abnormalities. We develop an image-level analysis of errors, which shows coarse location information exists in many apparently 'incorrect' location responses. Finally, we assess whether trials which imply detection of a target without localisation could be due to guessing. We demonstrate that for breast masses there is information that supports both detection and localisation of abnormalities, with better performance in images with low relative to high breast density. Our findings emphasise the need for breast density to be considered in screening reports and radiologist training. *Notification for the patient and clinician about breast density and potential cancer risk may have a significant positive effect on outcomes, such as the provision of more suitable imaging modalities, and an earlier cancer diagnosis.*

Background

As soon as we open our eyes, our visual system processes an enormous amount of information in a short space of time. Early findings showed that an exposure of 100ms is sufficient to extract the basic meaning of natural scenes (e.g., indoor versus outdoor; Potter, 1976). Using backward masking to precisely control for exposure times, others have shown that the distinction between natural scene categories at the superordinate

level (e.g., man-made versus natural) and basic level (e.g., coast versus city) can occur with presentation durations as short as 20ms (Greene & Oliva, 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007). Furthermore, when primed with a category (e.g., animal or truck), objects can be detected at brief durations (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). This fast visual processing has also been reported among those who are experienced in domain-specific tasks such as medical imaging (Evans, Georgian-Smith, Tambouret, Birdwell & Wolfe, 2013; Evans, Haygood, Cooper, Culpan & Wolfe, 2016; Kundel & Nodine, 1975; Nodine, Kundel, Mello-Thoms, Weinstein, Orel et al., 1999). Kundel & Nodine (1975) showed that when presented a chest radiograph for 200ms, radiologists could detect an abnormality with 70% accuracy. Kundel and colleagues (2008) have since shown that within 1 second of viewing a mammogram, experts fixate on 67% of breast cancers (Kundel, Nodine, Krupinski, Mello-Thoms, 2008). Furthermore, when shown briefly presented mammographic displays (250ms), radiologists can discriminate normal from abnormal at levels better than guessing (Evans et al., 2013; Evans et al., 2016). The evidence that observers can extract information with fast presentations from natural scenes (e.g., Potter, 1976; Thorpe et al., 2001), and medical images (e.g., Kundel & Nodine, 1975; Evans et al., 2013), suggests that the processing involved in early visual search is similar whether the display is a natural scene or a medical image, at least for experts.

Radiologists develop expertise in ‘visual search’ in such images over a period of years. It has been suggested that specialised training and ongoing experience leads to perceptual and cognitive ‘fine-tuning’ in the task of image interpretation (Nodine & Mello-Thoms, 2010). Maintaining such expertise requires interpreting high volumes of cases. For example, mammographic screening radiologists interpret more than 2000 cases per year (Rawashdeh, Lee, Bourne, Ryan, Pietrzyk, et al., 2013). It is possible that

expertise can be attributed to implicit learning and many hours of training and practice has allowed for the efficient guidance of attention to relevant regions in an image (Drew, Evans, K., Jacobson, & Wolfe, 2013). Evidence for expertise includes the findings that experienced radiologists outperform novices and trainee radiologists on tasks such as detecting an abnormality in brief images (Evans et al., 2013; Nodine et al., 1999), and in different patterns of eye movements between experts and novices. For example, Kundel and La Follette (1972) compared the visual scan patterns of expert breast radiologists with trainees interpreting mammograms and found that the experts fixated on lesions faster and concluded search earlier than the novices. Others have shown that experts fixate true abnormalities within 1-2 sec of image onset and most of their subsequent scanning is to confirm that there are no other lesions (Mello-Thoms, Hardesty, Sumkin, Ganott, Hakim et al., 2005). This follow-up takes about 5-10 seconds after initial fixation, after which a diagnostic decision is reached. There is an enormous amount of information that is processed in the first second of viewing a scene or image, so it is important that we understanding the cognitive underpinnings of early visual search.

Kundel and Nodine (1975) developed a model that describes two distinct processes leading to a diagnostic decision. The first glance supports a global, or holistic overview of the image, which indicates on a basic level whether the image deviates from a cognitive representation of a normal anatomical schema. The information extracted at this first stage is then proposed to constrain and guide search to the region of the image containing the abnormality (the second stage). For this to occur, the global signal must be informative about the location of the abnormality.

Recently an alternative perspective has been offered by Evans and colleagues (2013, 2016). They suggest an initial abnormal signal could act to alert a radiologist that *something* is abnormal but without containing location information. Rather than guiding

search to a location, this global signal then *changes the search strategy* to a more complete search for the abnormality. The initial signal could be supported by the rapid extraction of the summary statistics of the image, such as average orientation and size. In the basic vision literature, two stage models (e.g., Wolfe, Võ. Evans & Greene, 2011) describe an initial, non-selective pathway which, although limited in capacity, extracts summary statistics in parallel from the display. In the model, global processing occurs along this pathway. A second, selective pathway recognises one or a few objects at a time and requires selective attention. Together these pathways combine to support perception. Evans et al. (2013) and Evans et al. (2016) suggest that information via the non-selective pathway could alert a radiologist that something is abnormal, but the fine-grained detail, such as its location, only becomes available at the later selective stage.

Evans et al. (2013) compared the performance of radiologists and novices on the detection and localisation of abnormalities in mammograms. The stimuli were bilateral *(left and right breast)* mammograms where one side could contain subtle masses and architectural distortions that varied in size (10 to 48mm). Such pathologies are highly variable, and are difficult to detect and locate even by expert radiologists under free viewing conditions. As a result, these have the highest reported rate of false negatives (Knutzen & Gisvold, 1993). Despite these difficult images, Evans et al. (2013) found that radiologists (but not novices) could detect an abnormality above chance (Mean d' was ~ 0.7 for 250ms duration and up to ~1 for 2000ms duration, where d' of 0 is chance). For the combined detection and localisation task, images were displayed for 500ms. Following detection, the radiologists viewed a blank outline of the mammogram and were asked to localise by marking the abnormality with a mouse-click. Chance was determined by calculating the average percentage (across images) of overall tissue area lying within a predetermined region of abnormality. Although abnormalities could be detected by

radiologists above chance at 500ms, localisation performance was at chance. Evans et al. (2013) interpreted these results as evidence that the information extracted to support detection at brief durations does not contain location information, but is rather based on an overall 'gist' or holistic signal. In a subsequent paper, Evans et al. (2016) did another series of experiments using mammograms, replicating and extending their initial findings. In their second experiment, they presented radiologists a set of 120 single-sided (one breast) mammograms for 500ms and asked them to detect and then localise an abnormality. The unilateral mammograms either contained an abnormality (target-present), had no abnormality (target-absent), or was the contralateral breast from the target-present mammogram (no abnormality). In this experiment, mean d' for detection was 1.16 for the target-present/target-absent images, significantly above chance (0), whereas localisation accuracy was not significantly greater than that expected by chance (6%). They concluded that the radiologists could not localise a lesion despite detecting it. Further, they suggested that experienced radiologists could even make such judgments based on images from the contralateral (thus far normal) breast (remaining 40 images). Mean d' was 0.59 for detection of abnormality in the contralateral breast from a woman with signs of cancer in the other breast. This result is striking because the mammogram on which the judgement was based had no mass. These results provide intriguing hints that the information required for detection and that for localisation could be dissociable.

Evans et al. (2013) and Evans et al. (2016) interpret their results as reflecting a global signal of abnormality that lacks information about location of a specific mass. Indeed, the remarkable findings that a diagnosis could be made from the contralateral apparently-normal breast when the opposite side was abnormal might be explained by this interpretation. There are, however, some alternative interpretations that need to be carefully considered and ruled out. Frequentist statistics, used in these studies, cannot

distinguish between a true null (no effect exists) and a lack of sensitivity (an effect exists but is not detected). To interpret a null effect as evidence for there being no effect (in this case no localisation), we would need to use alternate statistics, such as a Bayes Factor (Dienes, 2011). Second, the summary statistics (e.g., average *d* prime) could be inadequate to answer the key questions. For d' values quite close to chance, artefacts or slight imprecisions in localisation for just a few images could be sufficient to drive performance to an apparently greater than chance level. For example, if participants are actually 'detecting' a distracting signal in the breast for a target present trial, the detection response would be correct but localisation would be incorrect (on the distractor). Similarly, if participants click just outside the lesion, this would be categorised as incorrect, which would lead to the erroneous inference that there was no localisation information. Finally, in a 2AFC (detection), there will always be some 'lucky guesses' that are correct. We need to consider the impact of these on the apparent dissociation between detection and localisation. These two studies by Evans and colleagues (2013) and (2016) raise important questions, but the challenge to the Kundel and Nodine (1975) model of radiologists' diagnostic decision-making rests heavily on the lack of information about the location of an abnormality. We need to go beyond the summary statistics and explore image level variability, precision of localisation responses and the potential influence of guesses to ascertain that there is truly detection without localisation.

The aims of the present study were to extend previous work by Evans and colleagues (2013; 2016) and explore in detail whether detection and localisation are dissociable. The claim that radiologists can detect the presence of an abnormality without knowing where it is has strong theoretical implications. Instead of the intuitive notion that the information in the first glance guides attention and the eyes towards the location of the potential abnormality, it implies a quite different process. Here, our first aim was to

see whether expert readers of mammograms viewing brief displays can extract location information when a mass is either obvious or subtle. Female breast tissue is highly variable in mammographic breast density (MBD: Li, Humphreys, Eriksson, Edgnen, Czene & Hall, 2013), which provides us with a natural variant for manipulating the salience of a mass. In the human population, 40% of women aged between 40-74 years have dense breasts (Sprague, Gangnon, Burt, Trentham-Dietz, Hampton, et al., 2014). Critically, as MBD increases there is a 4-6-fold increased risk of breast cancer (Boyd, Martin, Bronskill, Yaffe & Duric et al., 2010), and studies have shown that higher levels of MBD reduce radiologist sensitivity, thus limiting early detection of breast cancer (Al-Mousa, Ryan, Mello-Thoms & Brennan, 2014). For a radiologist, MBD increases the complexity of the image and could mask and/or distract from existing pathology. Our second aim was to *explore the effect of breast density (*which can make masses more difficult to see*)* on the type of information that can be extracted in a brief display. Finally, the distinction between theories rests heavily on the dissociation between detection and localisation of masses. Our third aim was therefore to develop methods that can test for evidence of this dissociation. To this end, we looked at the images in detail to explore the degree and source of localisation errors on apparent detection-correct trials, as well as considering the potential influence of 'lucky' guesses to 'detection without localisation' performance.

We investigate detection and localisation performance for a single mass in unilateral mammograms presented centrally for a brief duration and then masked. There is evidence of a bias to click directly in front of fixation (centre of the image) when the location is unknown (Buswell, 1935; Tatler, 2007). However, the mass location varied within the breast in our images, which minimises the influence of any such bias (i.e., a random central click is not likely to fall within the mass location). We presented two sets

of mammograms that varied on density (high density and low density) and mass presence. As half of the images contained a mass that would be difficult to detect, we used two durations (unique images in each): 250ms (within the timeframe others have considered to support gist-level information in medical images; Evans et al. (2013)) and 1000ms (presumably well beyond gist level of perception). The participants performed a detection and an 'exact click' localisation task similar to Evans and colleagues (2013). We had two conditions for our target-present stimuli, each containing a single mass: a difficult condition (50%) in which the mass was subtle due to level of breast density and an easy condition (50%) in which the mass was obvious. The difficult condition is comparable to those of Evans et al. (2013) and Evans et al. (2016). We predict that mass detection and localisation will be more accurate for mammograms with low density compared with those with high density at both experimental durations. We consider image variability, response imprecision and we use alternative analyses and a guessing correction to fully test for a dissociation between knowing an abnormality is present versus knowing where it is.

Method

Participants.

Twelve participants with experience in interpreting mammograms were recruited from BreastScreen New South Wales and local radiology practices (6 female, Average age = 54 years, *SD* = 13 years). We defined experts as having at least four years of experience and in their current practice reading at least 2000 mammographic cases per year (Rawashdeh et al., 2013). The BreastScreen doctors (n = 11) read > 3000 mammographic cases per year, but we did also include one breast physician who read > 1000 cases per year, as she had extensive experience (10 years). The average experience

reading mammograms of our participants was 22 years (*SD* = 13 years). All gave informed consent and reported normal or corrected-to-normal vision. The study was approved by the Macquarie University Human Research Ethics Committee (Medical Sciences).

Design, Stimuli and Apparatus

We used a Density (low, high) x Duration (250, 1000ms) within-subjects design. The stimuli were 160 full-field, de-identified, medio-lateral oblique digital breast mammograms obtained from the Dokuz Eylul Mammography Set (DEMS: Bulu, Alpkocak & Balci, 2013), which varied on target presence/absence, and high MBD/low MBD. Half the images (80) were normal and half contained a single mass previously diagnosed and coded according to the Breast Imaging and Reporting Data System (BIRADS: American College of Radiology: Breast Imaging Reporting and Data System Atlas. Reston, Va: © American College of Radiology, 2003). BIRADS is a standardised breast assessment tool developed for mammography that ranges from 0 to 6. In clinical practice, a radiologist assigns a BIRADS score to each image, which determines the next step in the diagnostic protocol. The 80 normal images had a previously assigned BIRADS code of one (no significant abnormality). The abnormal breast images consisted of BIRADS coded 2 (benign), 3 (probably benign), 4 (suspicious abnormality and biopsy recommended), 5 (highly suggestive of malignancy) and 6 (known pathological proven malignancy). The average size of the mass was 26.70 mm (*SD* = 13.23 mm) and the range was from 8 – 54 mm.

From this set, ten images were ‘cleaned’ using GraphicConverter (version 9.4). Image artifacts such as side markers and occasional dust speckles outside of the breast and large calcifications within the tissue were removed. One of the most challenging

aspects of studying radiologists and using medical images rather than using artificial stimuli is that the human body varies widely anatomically. Stimuli were selected that contained only a single mass (so those with a second lesion were excluded). Difficulty was manipulated by including two sets of mammograms (dense: high MBD and fatty: low MBD) where half of the mass images (40) and half of the normal images (40) had high MBD. The remaining images had low MBD (See Fig.1). Density was categorised on a dichotomous scale (low/high) by an experienced radiologist blind to the purpose of the study (M.B.) and one author with experience reading mammographic images (A.C.). These ratings were significantly correlated ($r = 0.9$, $p < .0001$).

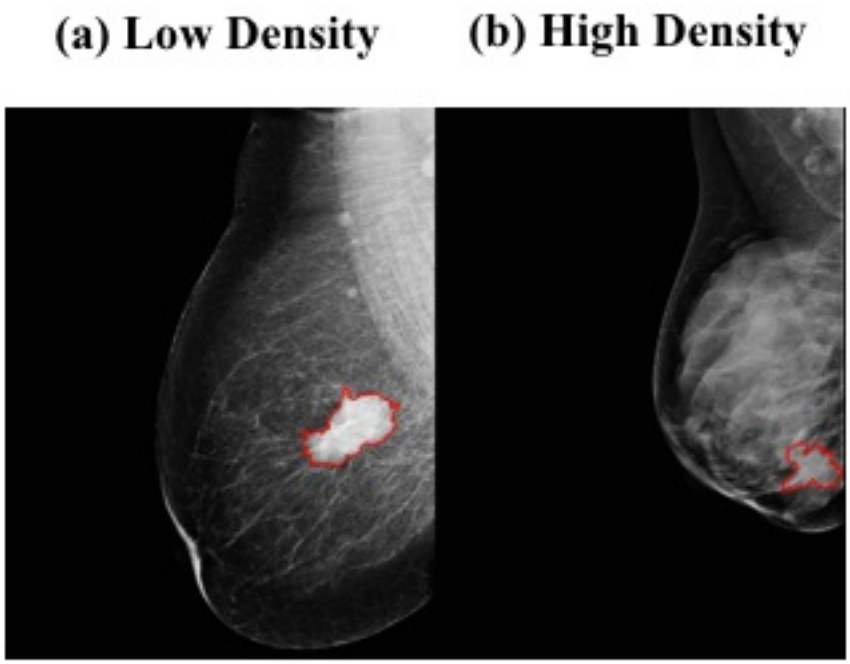


Fig. 1: Exemplars of target-present images. The red outline depicts the mass (and did not appear in the actual stimuli). (a) Low density breast that contains predominately fatty tissue, which is radio-translucent or black/grey. The higher contrast mass is easily seen; (b**)** High density breast that contains normal fibroglandular tissue resulting in a more

difficult search. The X-ray beam is attenuated by this tissue and appears radio-opaque or white on a mammogram.

The experiment was presented on a Macintosh MacBook Pro using MATLAB 2011B with the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). The stimuli were centred on a 1920 x 1080 resolution 24-inch, LG W2442PA, liquid-crystal display screen, refresh rate of 120Hz. The participants sat approximately 70cm away from the screen. The original resolution of the single mammograms was 4096 x 3328 or 3328 x 2560 pixels, which were downsized to 19° x 24° (18 out of 160) or 20° x 24° of visual angle. To validate our image categories and presentation durations, pilot data was collected from three radiologists at 250ms and 500ms durations two months prior to their participation in the experimental session. Previous studies which have used medical images have reported that a time-lapse of around 2 months between each session reduces the likelihood of recall (Berbaum, Krupinski, Schartz, Caldwell, Madsen et al., 2015). On the basis of these pilot data we increased the long duration condition to 1000ms.

Procedure

The experiment was conducted onsite at various metropolitan Sydney BreastScreen and radiology practice locations. We presented the stimuli at two presentation durations (250ms, 1000ms) in separate blocks, counterbalanced in order across participants. For each participant, the particular image presented in each duration was randomly selected without replacement. After four practice trials at 2000ms with feedback and a further six trials at the experimental durations (three at 250ms, three at 1000ms; blocked) with feedback, the radiologists viewed 160 trials without feedback. The radiologists were asked to detect ‘any mass that you would recommend for further

investigation'. Each trial began with a fixation point for 500ms, followed by a centrally-presented left medio-lateral oblique breast image. This was followed by a backward 1/f noise mask for 250ms after each stimulus presentation and a black screen asking the radiologists to categorise the mammogram using a key press as either 'normal' (left arrow key) or 'mass' (right arrow key), followed by a black screen with a grey mask of the breast (each unique mammogram was paired with its corresponding mask). The radiologists were asked to 'please click with the mouse the exact location where you saw a mass'. In the case of normal responses, they were asked to click anywhere on the display. There were 20 trials per condition (duration/target presence/density). Figure 2 shows the trial sequence. Participants began the next trial with a key press.

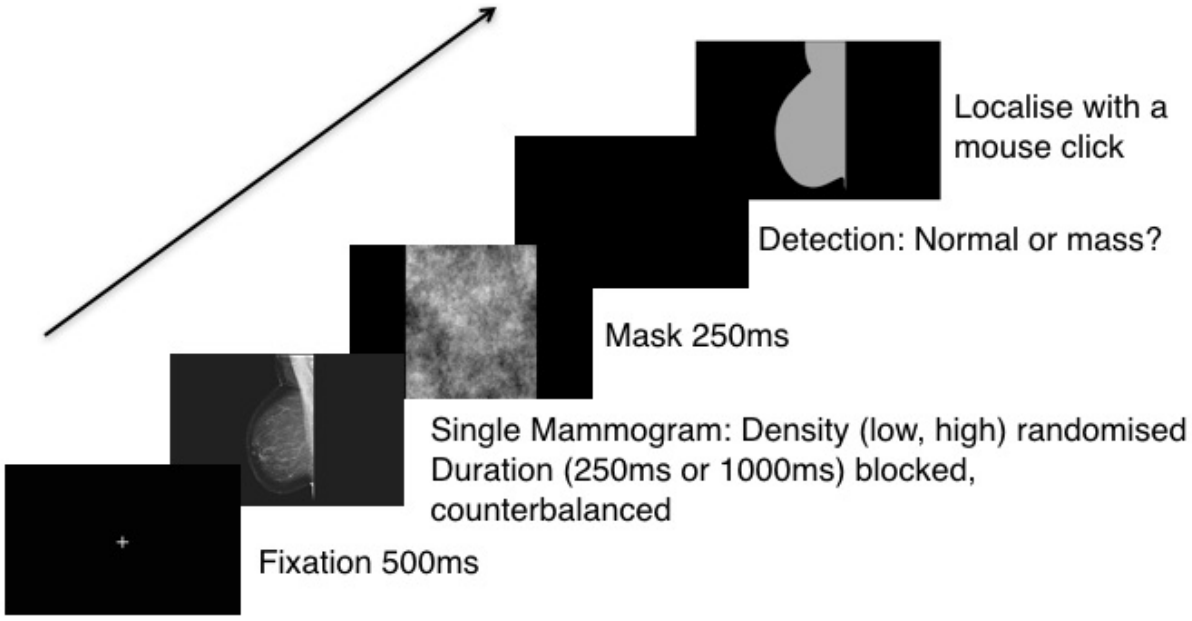


Fig. 2: Example trial for twelve radiologists who were asked first whether the image was normal or contained a mass, and then to use the mouse to indicate the location of the mass if present.

Analysis

Following the recommendations of Cumming (2012), we present Mean differences (M_{diff}) with 95% confidence intervals (CI), as well as a Cohen's *d* estimate of effect size corrected for small sample size, to assist in accurate interpretation of the effects. This latter measure, d_{unb}, represents an adjusted, unbiased Cohen's *d* standardised effect size applied to single sample t-tests where d_{unb} = (1 - 3 / (4*df - 1)) * *d* (Cumming, 2012).

Results

The aims of the experiment were to see whether expert readers of mammograms viewing brief displays (1) can extract location information; (2) are affected by breast density in the type of information that can be extracted; (3) show a dissociation between detection and localisation.

Detection accuracy: First, we calculated accuracy for target present and target absent trials to test whether the radiologists could detect a mass at these durations. Figure 3 shows performance on the detection task presented as accuracy for target present and absent trials separately (a, b) and sensitivity (c). Figure 3a shows better performance for low density images (more obvious masses) than the high density images (where the masses are more difficult to find even in free viewing). Accuracy also improves with duration. Figure 3b shows accuracy for the target absent trials. The radiologists appeared less accurate on target absent trials at the longer duration, showing they tended to make false alarms when given slightly more time to inspect the display.

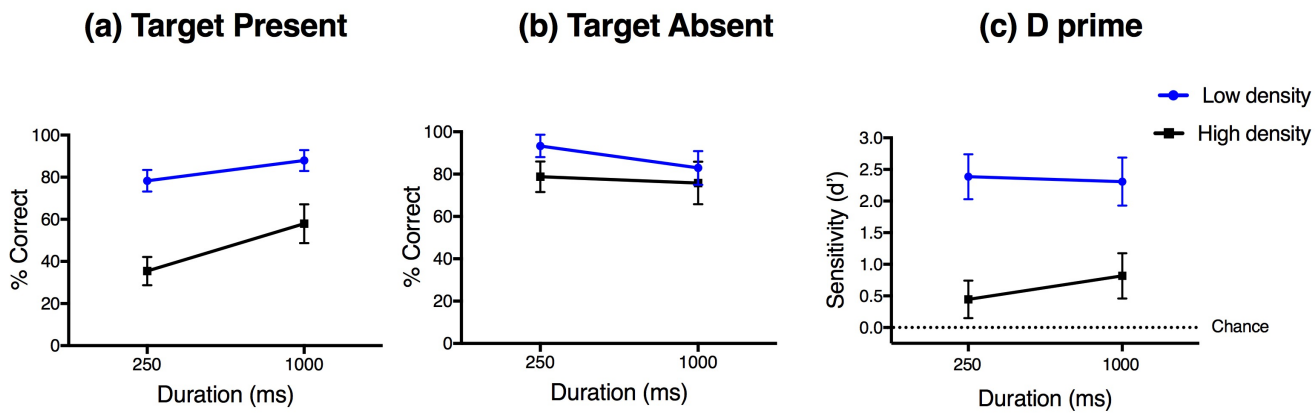


Fig. 3: Detection performance. (a) Average percentage correct on target present trials; (b) Average percentage correct on target absent trials; (c) Average sensitivity (d′) on the detection task. Error bars represent 95% confidence intervals.

Sensitivity (*d′*) was calculated as a function of abnormality present or absent. Higher *d′* indicates greater sensitivity: the higher the *d′*, the more accurately the radiologists responded to both target present and target absent trials (i.e., reported a mass when a mass was present *and* no mass when no mass was present). A *d′* of zero indicates there is no sensitivity and the participant is performing at chance (i.e., no better than guessing).

Figure 3c presents the sensitivity (*d′*) data. Single sample t-tests (Bonferonni adjusted, alpha = .0125) on average *d′* relative to 0 (chance) for each duration and density showed that radiologists do have information about the presence of the mass at both durations. Performance at 250ms for the low density condition was greater than chance

($t(11) = 14.97$, $p < .0001$, $M_{diff} = 2.39$, CI [2.03, 2.74], $d_{unb} = 5.69$) as was performance in the more difficult high density images ($t(11) = 3.3$, $p < .007$, $M_{diff} = .44$, CI [.15, .74], $d_{unb} = 1.3$). As one might expect, this was also the case at the longer duration of 1000ms, both for low density images ($t(11) = 13.38$, $p < .0001$, $M_{diff} = 2.31$, CI [1.93, 2.69], $d_{unb} = 5.09$) and high density images ($t(11) = 5.04$, $p < .0001$, $M_{diff} = .82$, CI [.46, 1.17], $d_{unb} = 1.92$). Although high density d' values reflect poorer performance than seen in free-viewing, where radiologists have d' values around 2.5–3.0 (D'Orsi Getty, Pickett, Sechopoulos, Newell et al., 2013), performance already approaches these levels for the low density images, even at 250ms (see Fig. 3c). These results suggest that when the mass is relatively easy to see (low density), diagnostic sensitivity in the first quarter of a second is already close to that of free-viewing.

As one would expect, we can see from Figure 3c that performance for the low density images is better than the high density images. This obvious pattern was confirmed by a repeated measures ANOVA with the factors of Density (low, high) x Duration (250, 1000) on the mean d' values. This showed a main effect of Density ($F(1, 11) = 133.51$, $p < .0001$, $\eta^2_p = .92$), no effect of Duration, ($F(1, 11) = .98$, $p = .344$) and no Density x Duration interaction ($F(1, 11) = 2.09$, $p = .18$). [1]

Localisation accuracy: Our key questions were first, whether there is localisation information when detection is correct, and second, how breast density influences localisation. Using the same method as Evans et al. (2013) and Evans et al. (2016, Experiment 2), we compared the location of the mouse click with the location of the actual mass and coded the response as either accurate (participant clicked on or within the boundaries of the mass) or not (any other location). We analysed trials where the participants were correct on detecting an abnormality at each exposure duration (i.e.,

[1] Individual observer data can be found in the supplementary materials.

correct detection target-present trials). We compared localisation performance to chance, calculated across the 80 target-present images as 4.4% (CI [3.02%, 5.75%]). This is the proportion of breast tissue that contains the mass relative to the proportion of total tissue; thus it represents the average number of possible random locations radiologists could select, taking into account the lesion and image size across all of the target-present images. Figure 4a shows the percentage of trials when the radiologists responded correctly on localisation task, when detection was correct, for low density (blue line) and high density (black line) at the two durations, compared with chance. Single sample t-tests (Bonferonni adjusted, alpha = .0125) showed that radiologists' localisation accuracy was significantly above chance (4.4%) for 250ms presentations of low density images ($t(11) = 12.9$, $p < .0001$, $M_{diff} = 30.18$, CI [25.03, 35.33], $d_{unb} = 4.9$) as well as for high density images ($t(11) = 3.74$, $p = .003$, $M_{diff} = 6.43$, CI [2.64, 10.22], $d_{unb} = 1.42$). The same pattern was evident at the longer duration of 1000ms for low ($t(11) = 13.9$, $p < .0001$, $M_{diff} = 50.6$, CI [42.59, 58.61], $d_{unb} = 5.28$) and high ($t(11) = 10.41$, $p < .0001$, $M_{diff} = 19.35$, CI [15.26, 23.44], $d_{unb} = 3.95$) density images.

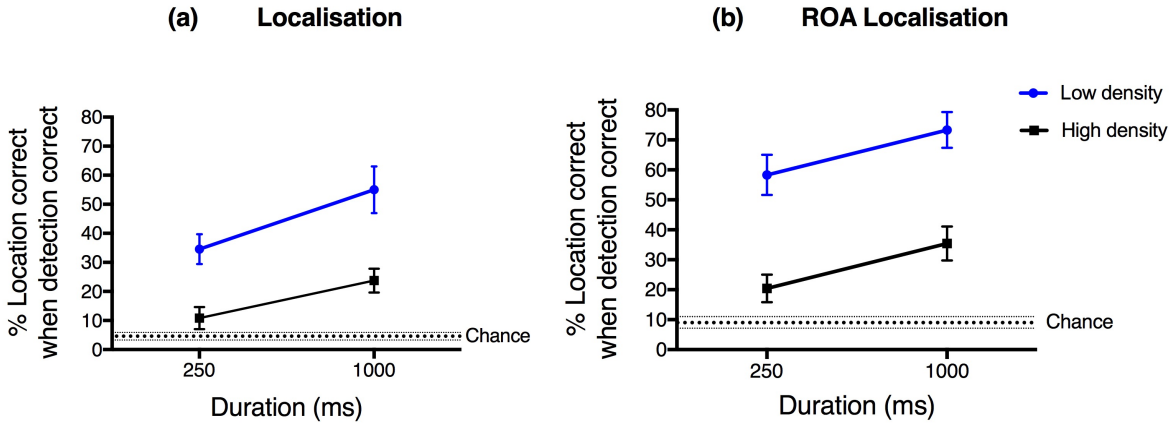


Fig 4: Detection and localisation results. (a) Average percentage correct on the localisation task for trials when detection was correct; (b) Average percentage correct on the localisation task when a region of acceptance (ROA) around the lesion is included. Chance is 4.4% and adjusted to 9.1% when including the ROA (dotted line) with 95% confidence intervals. Error bars represent 95% confidence intervals.

To investigate the effect of density on localisation (Fig. 4a), we conducted a repeated measures ANOVA with the factors of Density (low, high) x Duration (250, 1000) on the mean percentage localisation correct values from the correct detection target-present trials. Again in line with expectations, this showed a main effect of Density, with better localisation accuracy in the low than high density condition ($F(1, 11) = 114.07$, $p < .0001$, $\eta^2_p = .91$), a main effect for Duration, with better localisation accuracy at 1000ms than 250ms ($F(1,11) = 53.01$, $p < .0001$, $\eta^2_p = .83$), and no Density x Duration interaction ($F(1,11) = 2.17$, $p = .17$). These analyses show that radiologists were statistically above chance in localising the target on trials where they successfully detected a mass. However, as localisation performance is far from perfect, we have some trials on which detection occurred apparently without localisation information being

available. This could reflect a global signal as suggested in the previous literature (Evans et al., 2013) and to investigate this possibility thoroughly, we conducted several follow-up analyses.

Before concluding one has evidence of 'detection without localisation' (e.g., Evans et al., 2013; Evans et al., 2016), there are some important alternatives to be considered. First, we would like to note that before concluding anything from a null localisation effect (such as that of Evans and colleagues), we need to use statistics that can provide evidence of *no effect* (of localisation when there is detection) rather than just no evidence. Frequentist statistics do not allow for the interpretation of null effects – a *p* value greater than alpha merely informs us that we do not have evidence to reject the null hypothesis. To see whether there is evidence *for* the null hypothesis of *no localisation* information, we could instead calculate a Bayes Factor (BF). A BF < 1 indicates that the data support the null rather than the alternative hypothesis (BF < .33 provides strong evidence for the null), a BF ~ 1 indicates maximal insensitivity of the experimental evidence, whereas a BF >1 indicates the data support the alternative hypothesis (BF > 3 suggests strong evidence for the alternative) (Dienes, 2011). In our case, we do not have a null effect in any condition, but we can still calculate a Bayes equivalent of a single sample t-test compared to chance (4.4%) to illustrate the point: if we test just the difficult images that are comparable to those of Evans et al. (2013; 2016), we can see strong evidence for the alternative hypothesis that localisation information exists: For the high density condition at 250ms, the BF(12) = 14.73 and at the longer duration, 1000ms, BF(12) = 31052.09. Consistent with our frequentist statistics results, we conclude that the radiologists are localising targets better than chance in the high dense conditions.

Our second consideration is whether summary level statistics such as overall accuracy or sensitivity are adequate to address the 'detection without localisation'

question. In fact, one cannot be sure of 'detection without localisation' without examining the error trials carefully. A null localisation effect could, for example, be due to less precision in the localisation task than the detection task, due to the additional requirements rather than a true lack of localisation information. This could include decay in the visual short-term memory trace over time or motor error in clicking the precise location. If such factors influence the precision of the localisation responses, we should see localisation errors that nonetheless cluster around the correct region. Our radiologists were scored correct on localisation if the mouse-click occurred within or on the boundaries of the lesion, consistent with Evans et al. (2016) (Evans, personal communication, May, 2017). However, when we look at the incorrect localisation responses, we see that this does not accurately reflect the degree of localisation information. For example, in Figure 5a, many of the 'incorrect' responses suggest the participant had some information about location, rather than basing his or her response on an amorphous global signal of abnormality.

There is also inherent variability in real-world stimuli. Although we carefully selected images with only one true mass, and removed obvious artefacts (e.g., dust), the images have naturally-occurring variations in breast tissue. We need to examine the responses at an image level to assess whether such variance may have contributed to trials of apparent successful detection without accurate localisation. Figure 5b shows clearly an image where natural variability has contributed to three incorrect responses to a distractor in the breast (presumably in these cases, the radiologists were responding 'abnormality present' to this distractor, rather than the actual mass). The responses on these images suggest that apparent 'detection without localisation' may actually reflect coarse or less precise localisation, rather than no localisation, warranting image-level investigation.

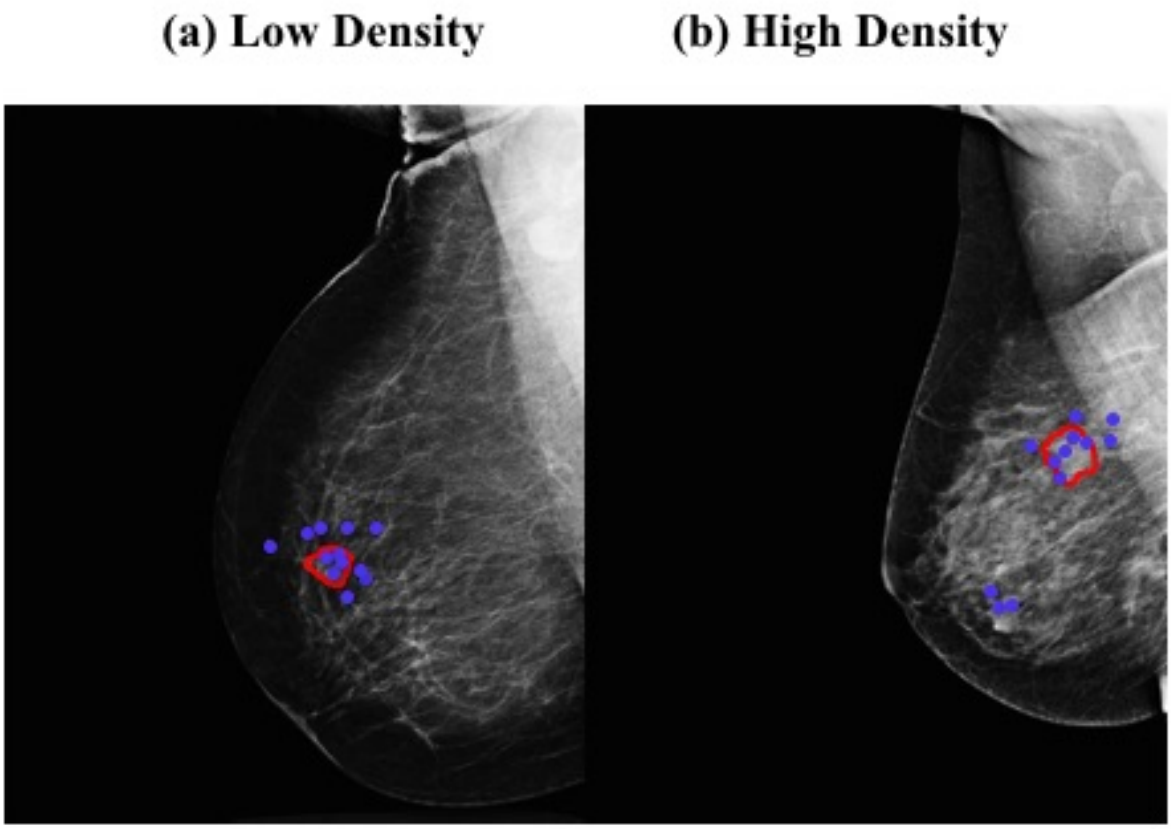


Fig 5: Exemplars from the target present stimuli set illustrating the mass (red outline, not shown in the experiment) and localisation responses of the 12 radiologists (blue) collapsed across duration. (a) Low density image showing precision errors. The blue mouse-clicks for localisation show that the 8 radiologists who were 'incorrect' on this image may have information about the location of the target; (b) High density image showing the effect of a naturally-occurring distractor. Three radiologists localised the distractor as the abnormality (note a further 4 'incorrect' responses are near the mass (red outline) but imprecise).

To quantify the degree to which such examples might influence our results, we conducted a post-hoc image analysis collapsed across participants for each duration. We calculated the distance between the response click and the mass (i.e., the degree of incorrect localisation). In academic radiology, a region of acceptance (ROA) for lesion localisation is determined by taking into account the size of the largest lesion (e.g., Haygood, Ryan, Brennan, Li, Marom, et al., 2014). Following this convention, we measured the radius of the largest mass in the image set (27mm) and added this value to the boundary values for all the target present images. Using this method, localisation is scored correct when a radiologist clicks within this ROA, allowing for a margin of response imprecision and reducing the 'tightness' of acceptance. We further examined the trials that were still incorrect to quantify the distance from the lesion boundary.

Figure 6 shows image level analysis for the localisation data on incorrect trials plotted as a function of distance (in pixels) from the closest boundary of the mass, collapsed across radiologists (Fig. 6a: 250ms; Fig. 6b: 1000ms). Trials on which the detection response was incorrect are not included (250ms: high density n=12, low density n= 1; 1000ms: high density n= 8; low density n= 0). Correct responses for localisation (when detection correct) would appear on the baseline and are also not included in the figure (250ms: high density n= 3, low density n= 8; 1000ms: high density n= 8; low density n= 8). The dashed red line represents the ROA plotted at 29 pixels. Figure 6 shows a considerable proportion of the clicks lie within this decision boundary and highlights how the variability within each image affected accuracy due to factors such as mass size and distractors.

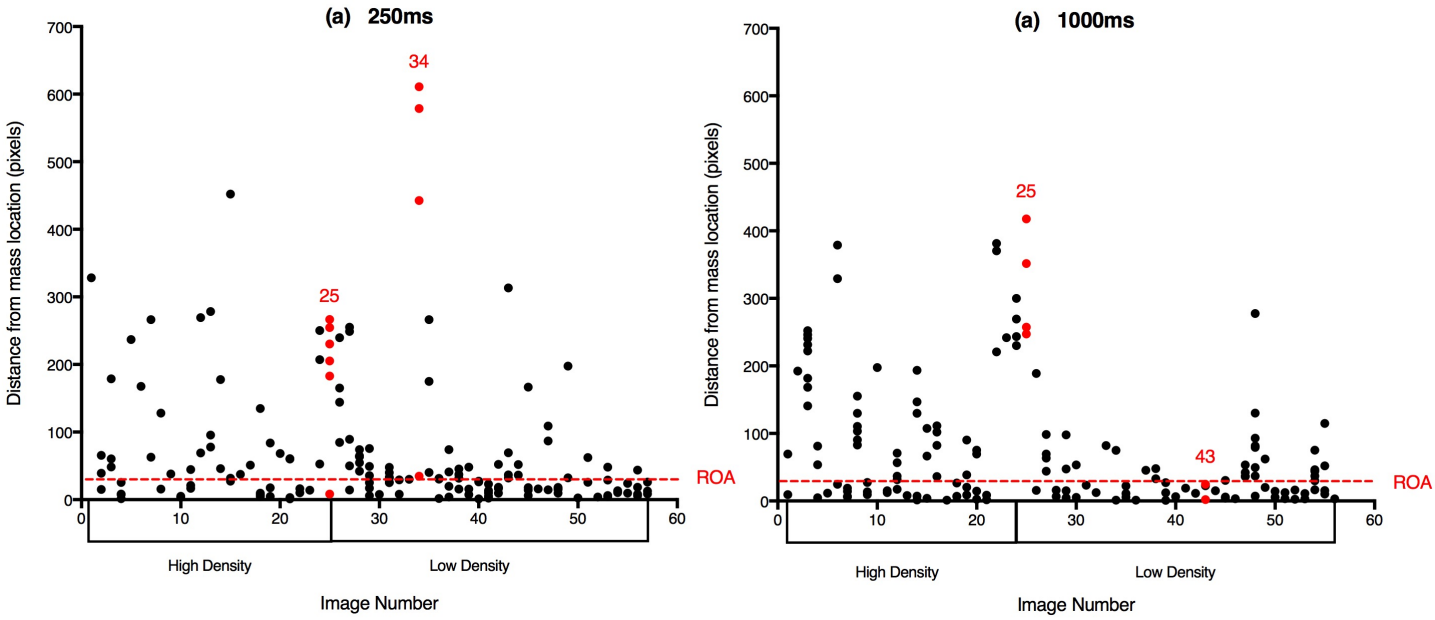


Fig 6: Localisation errors showing the distance between the localisation response and the mass for each image (detection correct target-present trials only). (a) 250ms duration; (b) 1000ms duration. The x-axis represents the images (divided by high and low density. Note: the image numbers are arbitrary for the purpose of the graph only). A correct score on localisation would score 0 (excluded from the figure). The y-axis is the distance (in pixels) from the mass border. The dashed red line represents the region of acceptance (ROA). Red numbers are data points in response to images with unusual characteristics: 25 (250ms) is the high density image presented in Figure 5b showing the mouse-clicks on a distractor. 34 is a low density image which contained a prominent lymph node in the axillary tail of the breast which appears to have captured 4 radiologists' attention; 25 (1000ms) is a low density image containing a small mass and 43 is the low density image presented in Figure 5a showing the cluster of mouse-clicks near the correct location.

Localisation accuracy including a ROA: We calculated percent correct for localisation trials with an ROA included in assessing localisation for target-present trials with correct detection responses. Figure 4b shows the percentage of trials in which ROA localisation was correct for low density (blue line) and high density (black line) images across both durations, compared with chance. ROA chance was calculated as 9.1%, adjusted to account for the increased proportion of tissue included in the ROA. The summary-level measures clearly indicate better accuracy for all conditions compared with the non-ROA data (Figure 4a), especially for the 250ms high density condition (ROA Mean = 20.42%; non-ROA Mean = 10.83%), demonstrating that the Evans et al. (2013) and Evans et al. (2016) method for calculating localisation may not adequately capture the degree to which location information is present.

This post-hoc analysis highlights the variability and challenges which exist when using real-world stimuli, and the importance of carefully examining the data from individual images rather than stopping at summary statistics. These findings suggest that the apparent lack of localisation on some trials where a mass was detected is, at least in part, driven by image variability, such as small masses in a proportionally large breast and normal tissue with salient features (distractors), and response imprecision. When we apply a more liberal localisation ROA, we see evidence that coarse localisation information exists, with a higher proportion of correct localisation responses even for the more difficult images.

We can also bin trials on which detection was correct according to their response profile to further examine the distribution of trial performance. Figure 7 shows the localisation data calculated using an ROA as a function of detection performance (collapsed across radiologists and images) for trials on which detection plus localisation were correct (blue bar), the additional localisation correct trials produced by including a

ROA (dark grey bar) and detection only trials on which localisation was incorrect (light grey bar).

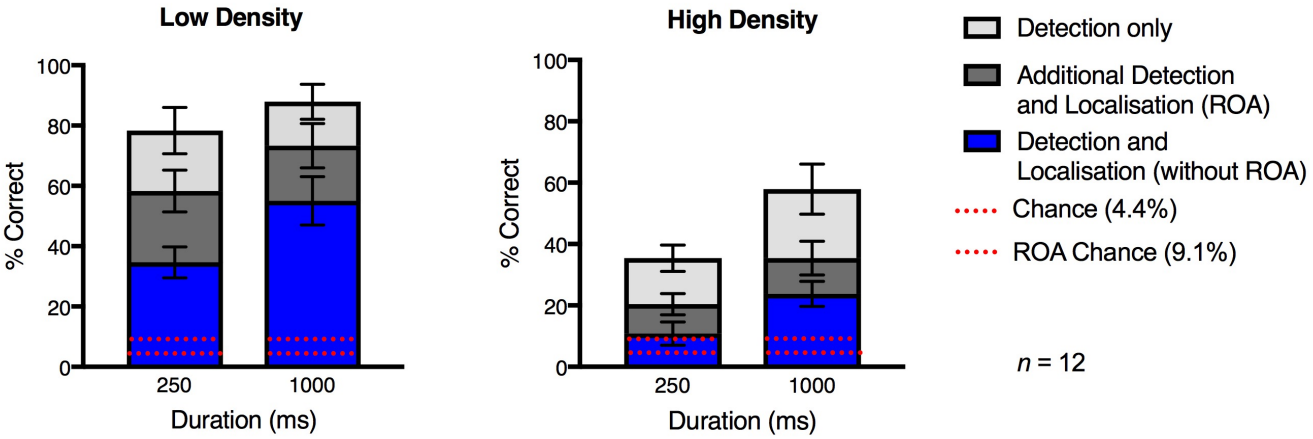


Fig 7: Percentage correct detection and localisation on target-present trials for low and high density mammograms plotted by duration (250ms, 1000ms). Data are separated by response accuracy: Detection and localisation correct (blue bar); the additional proportion of trials where localisation is correct when a ROA is included (dark grey bar); and detection correct/localisation incorrect (light grey bar). Error bars represent 95% confidence intervals.

In addition to the trials with evidence for coarse localisation or precise mis-localisation, Figure 6 shows some remaining trials on which localisation is clearly incorrect; these contribute to the light grey bars in Figure 7. These trials could be evidence for 'detection without localisation', which seems key to interpretations of radiologists using 'gist' or a global signal. However, there is one final consideration

before making such an interpretation: we need to be sure that the number of trials on which this occurs exceeds the rate at which such trials would occur simply from 'lucky' guesses. With any visual detection task, some proportion of trials will be correct by chance. A d' above chance shows more trials are correct than would be predicted by simply guessing, but if one wants to infer that there are trials in which there is 'detection without localisation', we need to calculate what proportion of these could be lucky correct detection guesses, followed by a localisation guess (which has less chance of being correct, recall chance in Evans et al. (2016) for localisation was ~ 6%).

We calculated a guessing probability using the method described in Howe and Webb (2014). They were interested in whether observers could ever 'sense' a change in a change blindness paradigm without knowing where the change was. In their method, one works out what proportion of correct detection trials (in their study, detection of a change) could be due to lucky guesses by creating a hypothetical observer who can only detect a change when it also knows what that change is (i.e., there is no true detection without localisation, therefore any such trials are due to correct guesses). Here, we used the same logic, a hypothetical observer who cannot detect a mass without also knowing where that mass is, to work out the proportion of trials on which correct detection combined with incorrect localisation could be due to lucky guesses. We can then compare actual performance with this prediction for each radiologist.

$$\textit{Calculated N (hypothetical observer)} = Q(Y-PA)/(1-P)$$

where Q = proportion of possible incorrect localisations; Y = number of target present trials on which the participant responded 'target present' (hits); A = actual number of target

present trials; and *P* = proportion of target absent trials on which the participant responded 'target present' (false alarms). (Note, there is no correction applied to an observer with no false alarms).

We calculate a guessing probability for the ROA localisation data, as this already takes into account any slight imprecisions in the localisation responses, giving the most accurate view of localisation information at a summary level. If the actual participants correctly indicated the presence of a mass in the absence of a correct location response more often than this hypothetical observer, this provides evidence for information about the presence of an abnormality without knowing where it is: 'detection without localisation'. Figure 8 shows the number of 'detection without localisation' trials from our data (dark grey bars) and the number of trials the hypothetical observer would 'guess' for all four conditions (light grey bars).

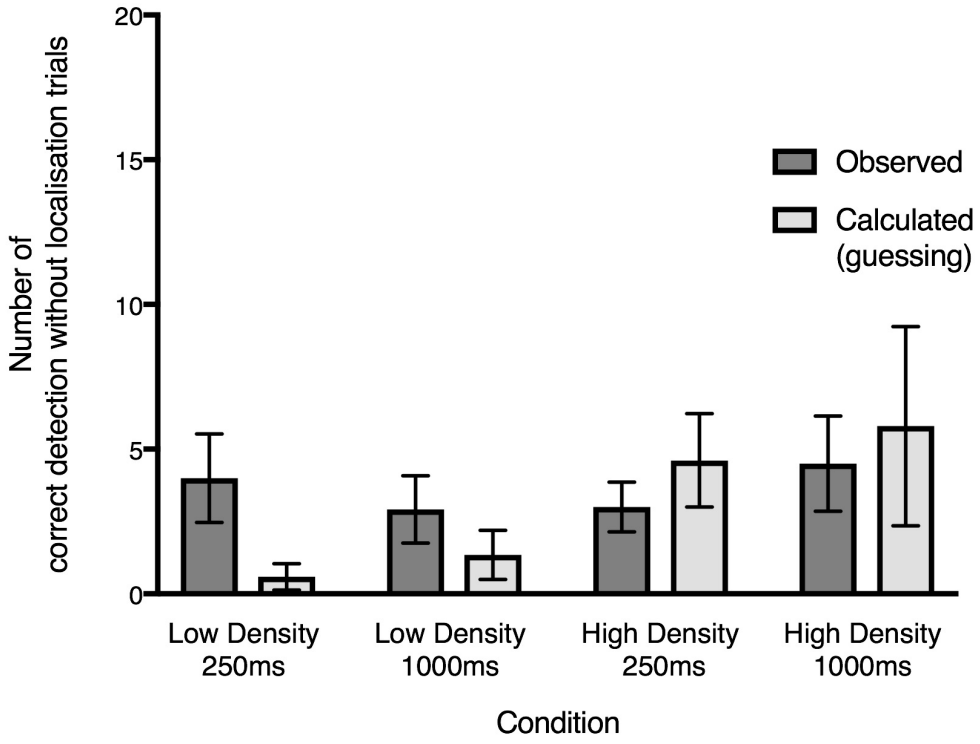


Fig. 8: The observed number of correct 'detection without localisation' trials (dark grey bars) compared to the number of calculated (guessing) trials for a hypothetical ideal observer (light grey bars) for low and high density mammograms plotted by duration (250ms, 1000ms). Error bars represent 95% confidence intervals.

From Figure 8, it is clear that there are only a small number of trials representing apparent 'detection without localisation', which makes statistical analysis unlikely to be reliable. However, even just from the graph one can see that only for the low density conditions is there any chance that there might be more detection without localisation trials than predicted by our hypothetical observer. From the image level analysis, these trials could reflect errors accounted for by response imprecision (e.g., large amount of breast tissue/small mass) and distractors. Recall that it is our high density condition that has images in which the mass is comparable in difficulty to Evans and colleagues (2013; 2016), making this the key condition. We have no evidence that for this high density condition the number of observed 'detection without localisation' trials is more than what would be predicted by 'lucky' guesses.

Discussion

The aim of this study was to examine the type of information that is available in the initial processing of a medical image (mammogram) by experienced radiologists, focusing on *detection* and *localisation* of potential abnormalities. We found radiologists were able to *detect* abnormalities at both durations (250ms, 1000ms) and density conditions (high, low), with a significant effect of duration. Overall summary statistics also supported the presence of *localisation* information, with the radiologists performing better than chance for both the 250ms and 1000ms durations, for the low and high density mammograms. Breast density affected performance in a predictable way, with better performance for low than high density images. As our key question related to a potential dissociation between detection and localisation, we carefully examined trials on which there seemed to be a dissociation. We suggest a number of factors that can lead to an underestimation of localisation information such as image variability, the precision of

localisation responses, and correct detection guesses. Overall, our data suggest that although it is possible that there may be a dissociation between detection and localisation on a small number of trials, particularly on easy trials (low density), there are other plausible explanations for the majority of such apparent dissociation trials.

Recent high-profile papers have concluded that radiologists can detect but not localise abnormalities in briefly presented mammograms (Evans et al., 2013; Evans et al., 2016). These papers suggest a different process to the previous theory that the information in the first glance guides experienced radiologists' attention and directs their eyes towards the location of the potential abnormality (Kundel & Nodine, 1975). Specifically, Evans et al. (2013) and Evans et al. (2016) proposed that the information extracted in the early signal is a global impression, which alerts the radiologist to the presence of an abnormality and then prompts a more thorough search, rather than guiding attention to the region of the abnormality directly. This alternative theory depends crucially on radiologists being able to detect masses in the *absence* of any information about location.

One of the key distinctions between the Evans et al. (2013; 2016) studies and our study is that they presented stimuli described as 'subtle masses and architectural distortions'. This might mean that there were a mix of both potentially localisable abnormalities (subtle masses) and abnormalities that do not have a well-defined location (architectural distortions, which do not contain a discrete mass in the parenchyma), or each category in a separate image. It would then make sense if there were no localisation signals as the abnormality may not have a well defined boundary or location. A global or gist signal also seems a plausible explanation for the other intriguing findings from these researchers in which radiologists are above chance in detecting abnormality in a patient when shown whole mammograms of a contralateral normal breast (Experiment 2) or only

a patch of a mammographic image that does not actually contain the mass (Evans et al., 2016, Experiment 4). In these cases, however, there is no mass to localise, making these findings less relevant to the question of whether a localisable mass can indeed be detected without being localised (although obviously pertinent to the idea that a global signal can be used to diagnose an abnormality). The evidence pertaining to this question comes from the experiments in which it seems there is a mix of pathology. It would therefore be interesting to know the proportion of these two types of breast pathology in the Evans et al. (2016) stimulus set, and how the location data break down by pathology. This would then allow a more accurate comparison with our own data.

With the Evans et al. (2013; 2016) studies as a whole, however, there may be influences other than gist leading to the results. Working with real-world images introduces many challenges, and with *d* prime values quite close to chance, we raise the concern that these data could be driven by a small number of images that contained additional artefacts. If these studies had any images like those illustrated in Figure 5, this could contribute to correct detection but incorrect location responses. Similarly, if the localisation responses cluster around the actual mass but not within the boundaries in some images, such as we found in our data (see Figure 6), this would also contribute to apparent detection without localisation. In the Evans et al. (2016) patch and contralateral breast experiments (2 and 4), coarse localisation cannot be an explanation, as there is no actual mass to localise. Thus, if there are any artefacts in images that drive detection above chance, this will appear to be dissociated from location (which is always chance). With only summary statistics showing *d* prime slightly (but significantly) above chance (Evans et al. (2013); Evans et al. (2016)), it is possible that the data interpreted as evidence for a global signal could be misleading.

Even when we use a conservative measure of localisation (click within the mass boundary), we were not able to replicate the findings of Evans et al. (2013) and Evans et al. (2016) that there are circumstances where radiologists can detect a mass above chance but not localise it. This could simply reflect that we were not at exactly the right durations to catch a dissociation due to variability in the experience of the participants, difficulty of the images, and other cross-experiment differences between our study and those of Evans and colleagues (2013; 2016). Another potential factor that could influence the difference between the studies is that our participants seem to be more experienced than those of Evans and colleagues (2013; 2016). This may be reason that we found localisation at a summary statistics level: our more experienced participants could extract information more rapidly and therefore processed the images in greater detail. For these previous studies to make the inference that there is *no localisation,* however, still requires an approach such as Bayes statistics, rather than standard frequentist statistics. Here, we have outlined the steps that seem crucial to be able to make an inference of dissociation between detection and localisation.

Although at the summary statistic level we did not replicate the lack of localisation information, we did find trials on which detection responses were correct but those for localisation were incorrect. We were therefore able to use these to investigate factors that might contribute to an apparent dissociation between detection and localisation. First, variability in the target-present images might be contributing misleading data to the summary statistics. Using real-world stimuli rather than typical laboratory visual search displays allows for high ecological validity, but the available images tend to be highly variable and it is difficult to control for factors such as co-existing variables (e.g., breast calcifications, target number and size and breast tissue type). Indeed, we identified images where there were clear clusters of incorrect

localisation corresponding to a specific visual feature in the image, suggesting the detection response was based on an incorrect identification (i.e., of the distracting feature). Second, we find evidence that coarse localisation information is often present in apparently incorrect responses. When we use a region of acceptance around the lesion, we see clusters of correct localisation responses surrounding the lesion. This suggests that task demands, such as having to hold the information through a detection response and subsequent location screen, may result in a loss of precision. Alternatively, it may be that the location information is only present at a coarse level in the first place (and is perfectly maintained). Finally, on trials where there is detection but incorrect localisation (by whatever definition one uses), it is important to consider the contribution of correct detection guesses. We used a method for estimating the effect correct guesses might have on the subsequent results. The key high density condition, which is most similar to that of Evans and colleagues (2013; 2016), gives no evidence for there being more 'detection without localisation' trials than can would be predicted to be lucky guesses. Thus, the pattern taken from a small number of trials suggest that in the difficult images, such as our set of high density mammograms, apparent 'detection without localisation' responses can be accounted for by 'lucky' guesses.

Our only evidence of an apparent dissociation between detection and localisation comes from the low density conditions. Intuitively, a salient mass seems most likely to have localisation information recorded, as there is a stronger bottom-up signal (much like a classic 'feature search'). Indeed, we do see overall better performance in the low density conditions compared with the high density conditions (although nowhere near 'pop-out' performance). Although our ROA takes into account coarse localisation information, it cannot account for image-level variability where a distractor may have been selected, or the potential decay of localisation information over time. Thus, while it

is possible that these potential ‘detection without localisation’ trials in the low density condition could reflect a global signal that is used to make a detection response, as proposed by Evans et al. (2013) and Evans et al. (2016), these trials could alternatively reflect the contribution of other factors to reducing localisation accuracy. Overall, such ‘detection without localisation’ occurred on a very small number of trials (~ 4), precluding statistical analysis, which means we have only the numerical difference to support any such inference. This means that for most of our stimuli, including those most similar to the previous studies, when the radiologists reported detecting a mass, they also had some information about where it was.

The proposal by Evans et al. (2013) and Evans et al. (2016) that radiologists use a global signal lacking in location information has important theoretical implications, as it identifies a very different mechanism from the Kundel and Nodine (1975) classic theory. Our results, however, demonstrate that successful detection of a mass in briefly presented mammograms is typically accompanied by information about location. This is more consistent with the Kundel and Nodine (1975) model: that the initial signal guides attention and eye movements to the lesion. To fully reconcile these distinctions, we need a study which investigates the presence (or lack thereof) of both global and localisable signals across three clearly defined conditions with different degrees of potential localisation (a salient mass, a subtle mass, or diffuse parenchymal change). We then need to ensure that the analyses are appropriate to the key question of whether any localisation information exists through a thorough image-level analysis.

Both detection and localisation performance decreased with increased breast density at fast presentations. These results are related to what we know about clutter in natural scenes and visual search in free viewing: increasing clutter or set size decreases performance (Adamo, Cain & Mitroff, 2015; Asher, Tolhurst, Troscianko, & Gilchrist,

2013; Rosenholtz, Li, Mansfield & Jin, 2005; Rosenholtz, Li & Nakano, 2007; Whitney & Levi, 2011; Wolfe, 1994). Fibroglandular tissue, which increases density on a mammogram, appears more radio-opaque than fat and may increase crowding and/or masking effects reducing performance in the denser mammograms. In the medical perception literature, there have been a number of studies that have investigated factors such as lesion subtlety, which may be dependent on the surrounding anatomical structures (e.g., Krupinski, 2005). Analogous to clutter interfering with performance in natural scenes, our results show similar effects in radiologists interpreting medical images.

These findings improve our understanding of how density can influence a radiologists' diagnostic decision and therefore have clinical relevance. Female breast tissue is highly variable with regards to mammographic breast density (MBD: Li et al., 2013) and high levels of breast density reduce radiologist sensitivity (see Al-Mousa, et al., 2014). It has been suggested that what radiologists perceive and thus report in the first second is critical (Mello-Thoms, 2009), that women with dense breasts make up almost a half of the population (Sprague et al., 2014), and that there is an increased risk of developing cancer in dense breasts (Boyd et al., 2010). Our results confirm that MBD has a negative impact on mass detection and localisation when radiologists are shown an image briefly. From a clinical viewpoint, we should inform women and their clinicians about their MBD levels, for appropriate and personalised care. For instance, in the case of a dense breast, further imaging modalities such as 3D mammography (digital breast tomosynthesis), ultrasound or magnetic resonance imaging will facilitate a definitive diagnosis. Although for almost half of the United States, density scoring is included (Slanetz, Freer & Birdwell, 2015), current breast screening reporting protocols in Australia do not include a mammographic density rating. Our data shows that high breast

density reduces the amount of information available in the first glance, suggesting reporting this information should be mandatory.

Conclusions

Here, we explored the degree to which information available in very brief presentations of medical images can support both detection and localisation of a mass in mammograms. Access to location information is crucial for guiding actions or further analysis (e.g., eye movements). We find a tight link between information supporting detection and localisation, using methods that allow a stronger test of the claim that detection of a mass can occur based on gist without knowledge of location. Although it is certainly possible that gist and the non-selective pathway of visual processing contribute to the detection of a non-localisable abnormality, our systematic examination of the factors that can result in apparent dissociation between detection and localisation demonstrates the importance of going beyond summary statistics when seeking to test this hypothesis. We emphasise the importance of considering factors such as stimulus variability, response imprecision, and participant guessing. Our results are consistent with Kundel & Nodine's (1975) model of radiologist visual search suggesting that the initial signal in a brief glance contains information that subsequently guides attention to the abnormality. Finally, we suggest the finding of reduced performance for dense mammograms illustrates the importance of reporting density information in the context of medical screening.

Abbreviations

ANOVA: Analysis of Variance; SD: Standard Deviation; MBD = Mammographic Breast Density

Acknowledgements:

For data collection, we acknowledge the support of BreastScreen NSW. We also thank Claudia Mello-Thoms for her initial advice and stimuli, Denise Moerel for her assistance with the image analysis, Ziba Gandomkar for assistance with the image database and Karla Evans for her feedback.

Ethics approval and consent to participate

Research was performed in accordance with the Declaration of Helsinki and was approved by Macquarie University Human Research Ethics Committee (Medical Sciences: Reference: 5201400567). Informed consent was obtained from all participants.

Duplicate publications

This research has previously been published as a conference abstract presented at the 57th Annual Meeting of the Psychonomic Society, Boston, USA. Expertise and the effect of density on detection and localisation in rapid presentation of mammograms. Carrigan, A.J., Wardle, S.G., & Rich, A.N. (2016, November).

References

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2015). Targets Need Their Own Personal Space: Effects of Clutter on Multiple-Target Search Accuracy. *Perception*, *44*(10), 1203–1214. http://doi.org/10.1177/0301006615594921.

Asher, M.F., Tolhurst, D.J., Troscianko, T., & Gilchrist, I.D. (2013). Regional effects of clutter on human target detection performance. *Journal of Vision*, *13*(5)*:25*, 1-15.

Al Mousa, D. A., Ryan, E. A., Mello-Thoms, C., & Brennan, P. C. (2014). What effect does mammographic breast density have on lesion detection in digital mammography? *Clinical radiology*, *69*(4), 333-341.

American College of Radiology (2013). Breast Imaging Reporting and Data System (BI-RADS) Atlas. Reston, VA, American College of Radiology.

Berbaum, K. S., Krupinski, E. A., Schartz, K. M., Caldwell, R. T., Madsen, M. T., Hur, S., Laroia, A.T., Thompson, B.H., Mullan, B.F., & Franken, E. A. (2015). Satisfaction of search in chest radiography 2015. *Academic radiology*, *22*(11), 1457-1465.

Boyd, N. F, Martin, L. J, Bronskill, M, Yaffe, M. J, Duric, N & Minkin S. (2010). Breast tissue composition and susceptibility to breast cancer. *J Natl Cancer Inst, 102*(16), 1224-1237.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Buswell, G. T. (1935). How people look at pictures: A study of the psychology of perception in art. *Chicago: University of Chicago Press.*

Bulu, H., Alpkocak, A., Balci, P. (2013). Uncertainty modelling for ontology-based mammography annotation with intelligent BI-RADS scoring. *Computers in Biology and Medicine, 43*(4), .301-311. DOI: 10.1016/j.compbiomed.2013.01.001.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*(3), 274-290.

Drew, T., Evans, K., Võ, M. L. H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics*, *33*(1), 263-274.

D'Orsi, C. J., Getty, D. J., Pickett, R. M., Sechopoulos, I., Newell, M. S., Gundry, K. R., Bates, S.R., Nishikawa, R.M., Sickles, E.A., Karellas, A., & D'Orsi, E. M. (2013). Stereoscopic digital mammography: improved specificity and reduced rate of recall in a prospective clinical trial. *Radiology*, *266*(1), 81-88.

Evans, K.K., Georgian-Smith, D., Tambouret, R., Birdwell, R.L., & Wolfe, J.M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin Review, 20*(6)*,* 1170 – 1175. doi: 10.3758/s13423-013-0459-3

Evans, K. K., Haygood, T. M., Cooper, J., Culpan, A.-M., & Wolfe, J. M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences*, *113*(37), 10292–10297. http://doi.org/10.1073/pnas.1606187113.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137-176.

Haygood, T. M., Ryan, J., Brennan, P. C., Li, S., Marom, E. M., McEntee, M. F., Itani, M., Evanoff, M., & Chakraborty, D. (2014). On the choice of acceptance radius in free-response observer performance studies. *The British journal of radiology, 86*(1021), pp.42313554-42313554.

Howe, P.D.L. & Webb, M.E. (2014). Detecting unidentified changes. *PloS ONE, 9*(11), e84490.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorisation and object interference. *Vision Research, 47*(26), 3286 – 3297.

Knutzen, A. M., & Gisvold, J. J. (1993). Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions. In *Mayo Clinic Proceedings*, *68*(5), 454-460.

Krupinski, E. A. (2005). Visual search of mammographic images: Influence of lesion subtlety. *Academic radiology*, *12*(8), 965-969.

Kundel, H. L., & La Follette Jr, P. S. (1972). Visual search patterns and experience with radiological images 1. *Radiology*, *103*(3), 523-528.

Kundel, H.L., & Nodine, C.F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*(3), 527-532.

Kundel, H. L., Nodine, C. F., Krupinski, E. A., & Mello-Thoms, C. (2008). Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic radiology*, *15*(7), 881-886.

Li, J., Humphreys, K., Eriksson, L., Edgnen, G., Czene, K., & Hall, P. (2013). Mammographic density reduction is a prognostic marker of response to adjuvant tamoxifen therapy in postmenopausal patients with breast cancer. *Journal of Clinical Oncology, 31*(18), 2249-2256.

Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., Stalder, S., & Maitz, G. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Academic Radiology, 12*(7)*,* 830-840.

Nodine, C.F., Kundel, H.L., Mello-Thoms, C., Weinstein, S.P., Orel, S.G., Sullivan, D.C., & Conant, E.F. (1999). How experience and training influence mammographic expertise. *Academic Radiology, 6*(10)*,* 575-585.

Nodine, C.F., & Mello-Thoms, C. (2010). *The handbook of medical image perception and techniques* (pp. 139-156). E. Samei, & E. A. Krupinski (Eds.). Cambridge: Cambridge University Press.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem. 2*(5), 509–522.

Rawashdeh, M. A., Lee, W. B., Bourne, R. M., Ryan, E. A., Pietrzyk, M. W., Reed, W.M., Heard, R.C., Black, D.A. & Brennan, P.C. (2013). Markers of good performance in mammography depend on number of Annual readings. *Radiology, 269*, 61 – 67.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17–17. http://doi.org/10.1167/7.2.17.

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005, April). Feature congestion: a measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 761-770). ACM.

Slanetz, P. J., Freer, P. E, & Birdwell, R. L. (2015). Breast-density legislation - practical considerations. *New England Journal of Medicine, 372*(7), 593-595.

Sprague, B. L., Gangnon, R. E., Burt, V., Trentham-Dietz, A., Hampton, J. M., Wellman, R. D., Kerlikowske, K., & Miglioretti, D. L. (2014). Prevalence of mammographically dense breasts in the United States. *Journal of the National Cancer Institute*, *106*(10), dju255

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 4–4. http://doi.org/10.1167/7.14.4

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 382*(6582), 520 – 522.

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artefactual objects. *Perception, 30*(6), 655-688.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, *15*(4), 160-168.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in cognitive sciences*, *15*(2), 77-84.

Supplementary Materials

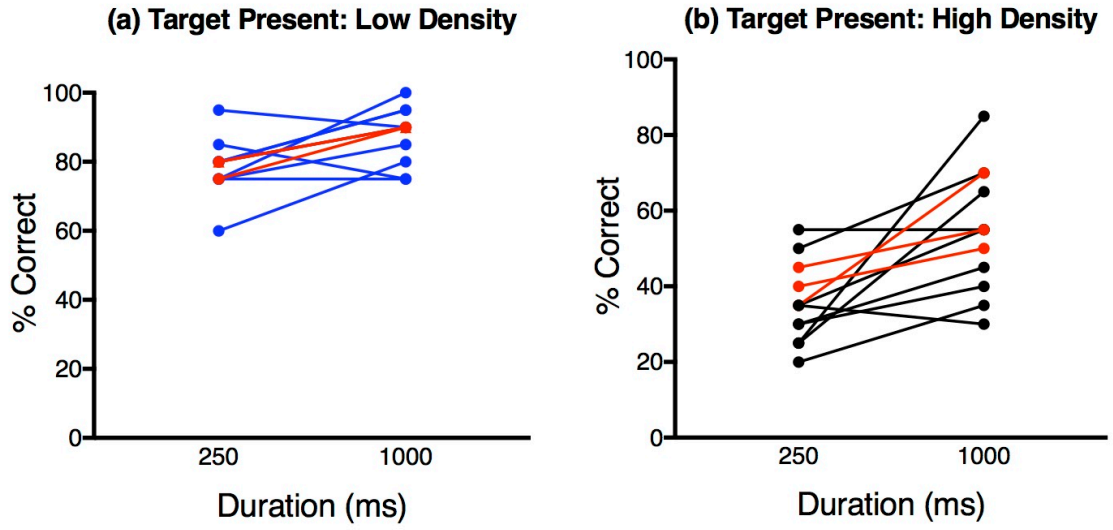


Fig.S1: Detection accuracy: percentage correct for individual radiologists on target present trials for (a) Low density and (b) High density mammograms on the detection task. The three radiologists that had piloted the experiment previously are illustrated in red.

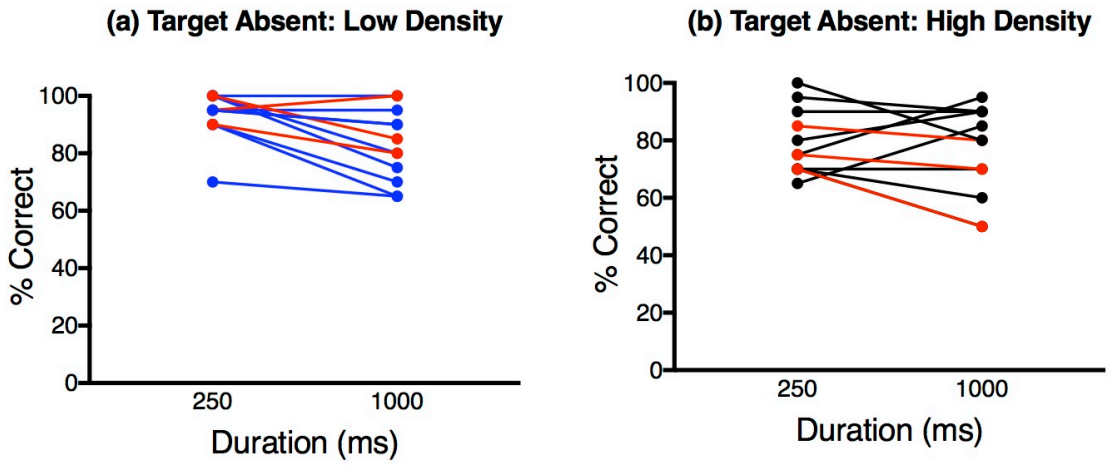


Fig. S2: Detection accuracy: percentage correct for individual radiologists on target absent trials for (a) Low density and (b) High density mammograms on the detection task. The three radiologists that had piloted the experiment previously are illustrated in red.

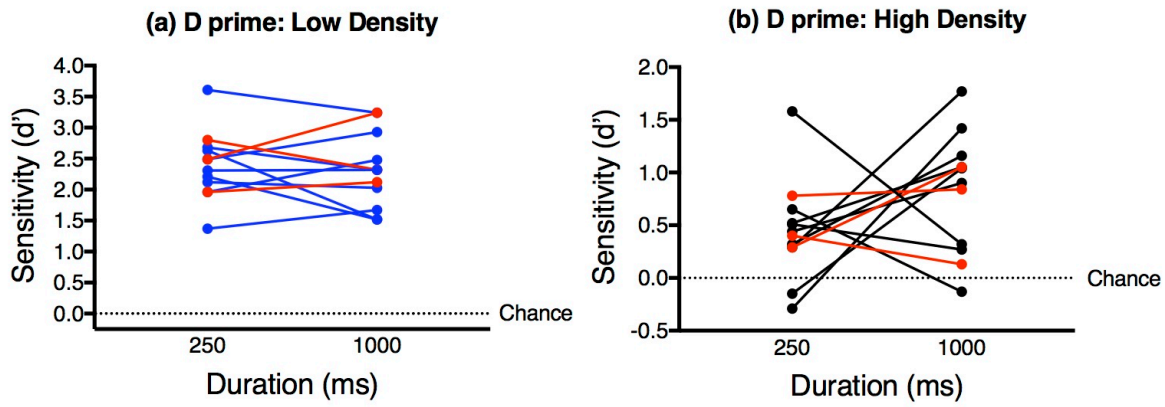


Fig.S3: Detection performance: sensitivity (d′) for individual radiologists (a) Low density and (b) High density mammograms. The three radiologists that had piloted the experiment previously are illustrated in red.

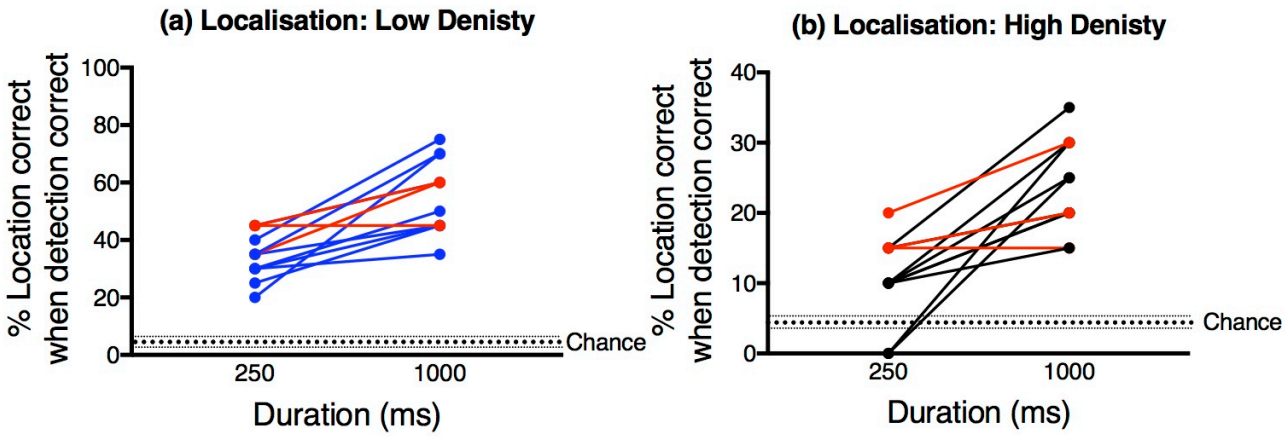


Fig. S4: Detection and localisation results: percentage correct on the localisation task for individual radiologists on trials when detection was correct for (a) Low density and (b) High density mammograms. The three radiologists that had piloted the experiment previously are illustrated in red. Chance is 4.4% and adjusted to 9.1% when including the ROA (dotted line) with 95% confidence intervals.

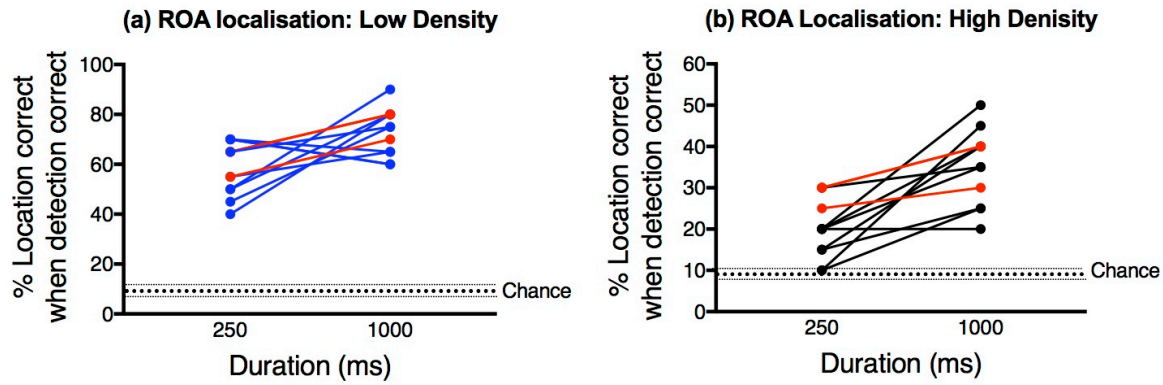
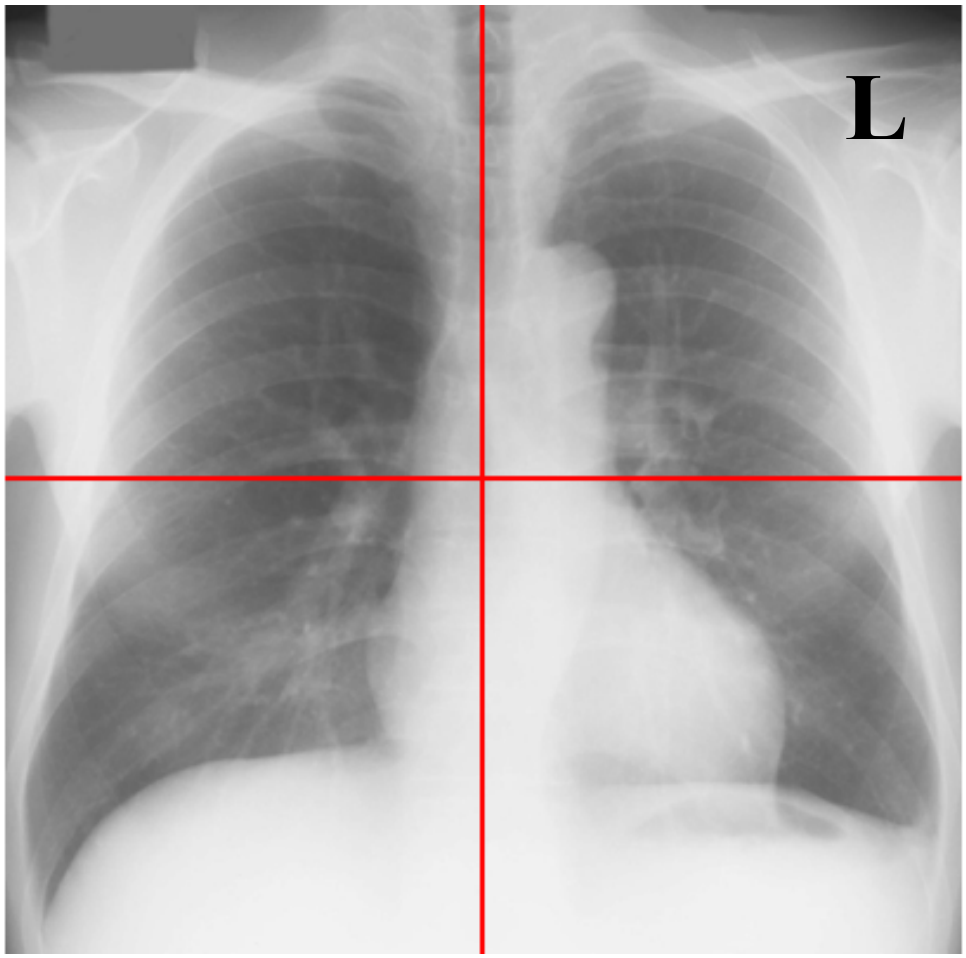


Fig. S5: Detection and localisation results: percentage correct on the localisation task when a region of acceptance (ROA) around the lesion is included for individual radiologists for (a) Low density and (b) High density mammograms. The three radiologists that had piloted the experiment previously are illustrated in red. Chance is 4.4% and adjusted to 9.1% when including the ROA (dotted line) with 95% confidence intervals.

Appendix B: Nodule Priors Questionnaire

Participant code _____

This is a posterior-anterior chest radiograph. Please mark 1-4 where you think the likely location for a single pulmonary nodule would occur, with 1 = most likely, 2 = likely, 3 = less likely, 4 = least likely.



Do you know the frequencies of nodules in different areas?

Yes/No

Thank you for your time ☺

An appendix of this thesis has been removed as it may contain sensitive/confidential content