

Published in final edited form as:

*Psychon Bull Rev.* 2013 December ; 20(6): . doi:10.3758/s13423-013-0459-3.

## The Gist of the Abnormal: Above chance medical decision making in the blink of an eye

Karl K. Evans<sup>1,2</sup>, Diane Georgian-Smith<sup>1,2</sup>, Rosemary Tambouret<sup>2,3</sup>, Robyn L. Birdwell<sup>1,2</sup>, and Jeremy M. Wolfe<sup>1,2</sup>

<sup>1</sup>Brigham and Women's Hospital

<sup>2</sup>Harvard Medical School

<sup>3</sup>Massachusetts General Hospital

### Abstract

Very fast extraction of global structural and statistical regularities allows us to access the “gist” – the basic meaning - of real world images in as little as 20 milliseconds. Gist processing is central to efficient assessment and orienting in complex environments. It is probable that this ability is based on our extensive experience with the regularities of the natural world. If that is so, would experts develop an ability to extract the gist from artificial stimuli (e.g. medical images) with which they have extensive visual experience? Anecdotally, experts report some ability to categorize images as normal or abnormal before actually finding an abnormality. We tested the reality of this perception in two expert populations; radiologists and cytologists. Observers viewed brief (250–2000 milliseconds) presentations of medical images. The presence of abnormality was randomized across trials. The task was to rate the abnormality of an image on a 0–100 analog scale and then to attempt to localize that abnormality on a subsequent screen showing only the outline of the image. Both groups of experts had above chance performance for detecting subtle abnormalities at all stimulus durations (cytologists  $D' \sim 1.2$  and radiologists  $D' \sim 1$ ) while non-expert control groups did not differ from chance ( $D' \sim 0.23$ ,  $D' \sim 0.25$ ). Further, the expert's ability to localize these abnormalities was at chance levels suggesting that categorization was based on a global signal and not on fortuitous attention to a localized target. It is possible that this global signal could be exploited to improve clinical performance.

---

Our visual world is very rich and complex, providing us with more information than our visual system can handle. Nevertheless, in spite of limitations in visual processing, we are still able to perceive significant information about a scene after a fraction of a second's exposure to it. An exposure on the order of 100 msec enables us to assess the general meaning or “gist” of a completely novel scene (Potter & Faulconer, 1975; Intraub, 1981). A 20 milliseconds masked exposure is enough to categorize the basic (e.g., lake vs. forest) or superordinate (e.g., natural vs. urban) level of a scene with above chance accuracy (Greene & Oliva, 2009; Joubert, Rousselet, Fize & Fabre-Thorpe, 2007). If primed with a category (e.g. animal), observers are above chance at detection of large objects (Thorpe, Fize & Marlot, 1996; VanRullen & Thorpe, 2001), even when focused attention is occupied with another foveal task (Li, VanRullen, Koch & Perona, 2002). In fact, observers are capable of rapidly extracting information about *multiple* categories even if they do not know the target category (animal, beach, mountain, etc) in advance (Evans, Horowitz & Wolfe, 2011). These abilities appear to be based on interpretation of global properties and image statistics,

based on our experience with the regularities in the natural world (Evans, Horowitz & Wolfe, 2011; Wolfe, Vo, Evans & Greene, 2011).

Medical experts, performing complex and much more artificial perceptual tasks, sometimes report that they feel as if they can categorize an image as normal or abnormal in a single glance. There are reports from medical image perception literature of radiologists detecting lesions in chest radiographs and mammograms at above chance levels with only a quarter of a second glimpse of the image (Kundel & Nodine, 1975; Carmody, Nodine & Kundel, 1981; Oestmann, Greene, Kushner, Bourgouin, Linetsky & Llewellyn, 1988; Mugglestone, Gale, Cowley & Wilson, 1995). These studies have typically been interpreted in the context of “the hypothesis that visual search begins with a global response that establishes content, detects gross deviations from normal, and organizes subsequent foveal checking fixations.” Of course, there can be ‘gross deviations’ that guide the deployment of attention. We hypothesized that experts can also sense a global signal, akin to the signals that allow for rapid natural scene categorization. This signal would not necessarily ‘organize subsequent fixations’ and find suspicious regions as hypothesized by Kundel and colleagues but would rather contribute to a conviction that a subsequent search would uncover an abnormality. We suggest that the global signal (i.e. gist) is an implicit extraction of statistics across the whole image allowing for categorization of the image that does not support precision object recognition within the image or constrain future eye movements.

To evaluate this hypothesis, we tested two sets of medical experts on their ability to extract the gist of the “abnormal” using briefly presented images from their domain of expertise. Fifty-five radiologists were presented with 100 trials of craniocaudal or mediolateral oblique x-ray views of both breasts (Figure 1a). Thirty-eight cytologists saw 120 Pap test images (micrographs of many cervical cells) (Figure 1b). Exposures were from 250 to 2000 milliseconds. Presence of abnormality and duration were randomized across trials. Observers rated the abnormality of an image on a 0–100 analog scale. Half of the cases were verified as not having any abnormality. The other half had various subtle abnormalities. For each expert group we tested a control group of naïve observers who had no significant experience with images of this sort.

## Materials and Methods

### Participants

We tested 55 radiologists (32 female, average age 56), 38 cytologists (22 female, average age 51), and 60 non-expert control observers (36 female, average age 30). All had normal or corrected-to-normal vision and all gave informed consent. All medical expert recruits were actively engaged in the daily practice of laboratory cervical cytology and radiology screening and had at least 5 years of experience. The group of radiologist experts had an average of 18 years of experience and a range of estimated 1,000–15,000 cases diagnosed a year. The group of cytologists had an average of 21 years of experience and a range of estimated 1,500–18,000 cases diagnosed a year. Naïve observers were recruited from the greater Boston area, had no medical training and were randomly assigned to view either cytology or mammography images. They performed the same task as experts after a short tutorial.

### Stimuli and Procedure

All observers viewed the images for a very brief time, 250 to 2000 milliseconds. Each observer saw a mix of two durations. In the first experiment, twenty radiologists saw images for 500 and 100 millisecond, thirteen for 250 and 2000 millisecond, seven for 750 and 200 millisecond. In Experiment Two, fifteen radiologists saw a 100 images all for 500

milliseconds. The control group of naïve observers saw images for 250 and 1000 milliseconds. In Experiment Three, both the expert group of cytologists and the control group saw the images for 250 and 1000 milliseconds. After 10 practice trials, both the expert group of radiologists and their control group completed 100 trials in which they viewed craniocaudal or mediolateral oblique views of x-ray images of breasts (mammograms). The expert group of cytologists and their control group completed 120 trials viewing Pap test images (micrographs) of many cervical cells. All three experiments were conducted on Dell Studio computer and the images were displayed on 19-inch, liquid-crystal color display screen at a viewing distance of 53 cm. The resolution of the monitor was  $1920 \times 1200$  pixels with usable intensity range of 2–260 candela per square meters and contrast ratio of 188:1. Half of the images were normal and half showed cancerous abnormalities. The abnormalities in the mammograms were subtle masses and architectural distortions while the abnormalities in the micrographs were both low-grade squamous intraepithelial lesions (LSIL) and high-grade squamous intraepithelial lesions (HSIL). In addition, after viewing, but before rating the images, 15 radiologists in second experiment and all of the cytologists were asked to localize the abnormalities by clicking on the display screen where they thought they saw an abnormality. The micrographs were followed by a blank, black screen and the mammograms by an outline of the breasts on which the observers were asked to make their best guess about the location of the abnormalities.

### Data analysis

We measured performance in terms of  $d'$  derived from the confidence ratings that ranged from 0 to 100. We used the rating of 50 as neutral divider of ratings to convert to binary responses of “YES, there is an abnormality”, or “NO, there is no abnormality”, responses. We adopted this measure for two reasons. First,  $d'$  is theoretically independent of an observer's bias to respond “yes” or “no”. Second, it is normally distributed, unlike accuracy, which makes it more suitable for standard parametric statistics. In addition to  $d'$ , we calculated a related measure derived from the ratings; the area under curve (AUC) of the receiver operating curve (ROC).

Localization performance was measured by determining the percentage of observer's clicks inside predetermined regions of interest delineating the abnormalities. Localization was assessed for trials on which observers correctly rated the image as abnormal. Chance levels for localization performance were determined by calculating the average percentage of overall tissue area covered by regions of interest.

### Results and Discussion

Both groups of experts had above chance performance for detecting subtle abnormalities at all stimulus durations (cytologists  $D' \sim 1.2$  and radiologists  $D' \sim 1$ ; Figure 2a & 3a). Neither of the control groups achieved significantly better than chance performance at short durations of 250 milliseconds and were considerably worse than experts at 1000 millisecond (Figures 2b & 3b). For the radiologists, tested at 5 different exposure durations, t-tests on  $d'$  show that performance was significantly above chance at each exposure duration (250 ms,  $t(19)=6.82$ ,  $p<0.0001$ ,  $AUC=0.64$ ; 500 ms,  $t(19)=11.28$ ,  $p<0.0001$ ,  $AUC=0.65$ ; 750 ms,  $t(7)=10.11$ ,  $p<0.0001$ ,  $AUC=0.65$ ; 1000 ms,  $t(19)=9.79$ ,  $p<0.0001$ ,  $AUC=0.66$ ; 2000 ms,  $t(12)=9.86$ ,  $p<0.0001$ ,  $AUC=0.72$ ) (Figure 2). Eye-movements (e.g. to each breast) were not required for this rapid gist extraction. The 250 millisecond condition does not permit volitional eye movements to each breast but was not significantly worse than 1000 millisecond exposure ( $t(19)= 1.8137$ ,  $p=0.0856$ ). Cytologists showed a similar pattern of results with above chance performance at both exposure durations (250 ms,  $t(37)=16.22$ ,

$p < 0.0001$ ,  $AUC = 0.71$  and  $1000\text{ ms}$ ,  $t(37) = 16.37$ ,  $p < 0.0001$ ,  $AUC = 0.77$ ) (Figure 3a & 3b). In this case, there was significant improvement at longer durations ( $t(37) = 4.42$ ,  $p < 0.0001$ ).

After rating a briefly exposed stimulus, all cytology experts and a smaller group of 15 radiologists, in a second experiment, were asked to localize abnormalities on a screen showing only the outline of the image (Figure 4a). For a localization to be deemed “correct”, it needed to fall into a region of interest (ROI), delineating the abnormality. These ROI’s were defined for each image by one of the authors: DGS for mammography, RT for cytology. We calculated the percentage of correctly localized abnormalities in respect to the overall number of abnormalities. Chance level was defined as percentage of the image falling in the abnormal region – the percentage that might be achieved by making random localizations. Localization performance for both groups was very poor, not significantly different than chance (cytologists: 16% correct localization; radiologists 15%; Figure 2c & 3b). Interestingly, the localization performance did not improve as the confidence rating increased staying flat for both expert groups and for both exposure durations in the cytology group.

This result should not be mistaken for one of those claims that we can or should make important decisions in the blink of an eye. No one would suggest performing cancer screening in 250 msec with a  $d'$  of  $\sim 1.0$ , given that the performance of expert radiologists is  $d' = 2.5\text{--}3.0$ . This result does not show that assessment is over in a blink of an eye. Rather it shows that, with the correct training, experts can have a global impression of the normality or abnormality of a medical image. That impression appears to be based on a global signal that, by itself, is not sufficient to localize the target. Mack and Palmeri (Mack & Palmeri, 2010) draw similar conclusions about real scene categorization based. They find that a computational model solely based on global scene statistics can explain the consistent-object advantage in rapid scene categorization with out any explicit rapid object recognition within a scene. Thus global image statistics are sufficient to differentiate scene with objects that are consistent or inconsistent with overall scene context. We believe that this is a trained, specialization of normal gist processing. Experience with the world has taught all of us that *this* set of image statistics is typically associated with, for example, an urban street, while *that* set of image statistics is associated with farmland. With specific training, an expert radiologist or cytologist learns the statistical regularities that distinguish normal from abnormal in the images in their realm of expertise. The ability to have a feeling that something is amiss and yet not know where to find it is akin to what Rensink called “mindsight” where observers consciously sense that a change occurred but still have no visual experience of that change (Rensink, 2004). However the method establishing mindsight by looking at the time difference between the observers “sense” response as opposed to “saw” response in a change detection task has been put into question (Simons, Nevarez & Boot, 2005). Simons and colleagues argue that the findings Rensink attributes to mindsight could be explained as a result of verification process rather a shift from an initial liberal decision criterion to a more conservative criterion.

The utility of this talent should not be oversold, it would be unwise to declare an image ‘normal’ on the basis of this global signal. However, devoting extra scrutiny to images that *feel* abnormal might improve performance. Moreover, if the signal could be identified by a computer, it could be used as a novel form of computer aided detection (CAD). Normal mammography CAD markings are used to indicate possible locations of targets. A global CAD would simply be a warning that this is the type of case that has an elevated chance of giving rise to a positive finding (see Hope et al. 2013).

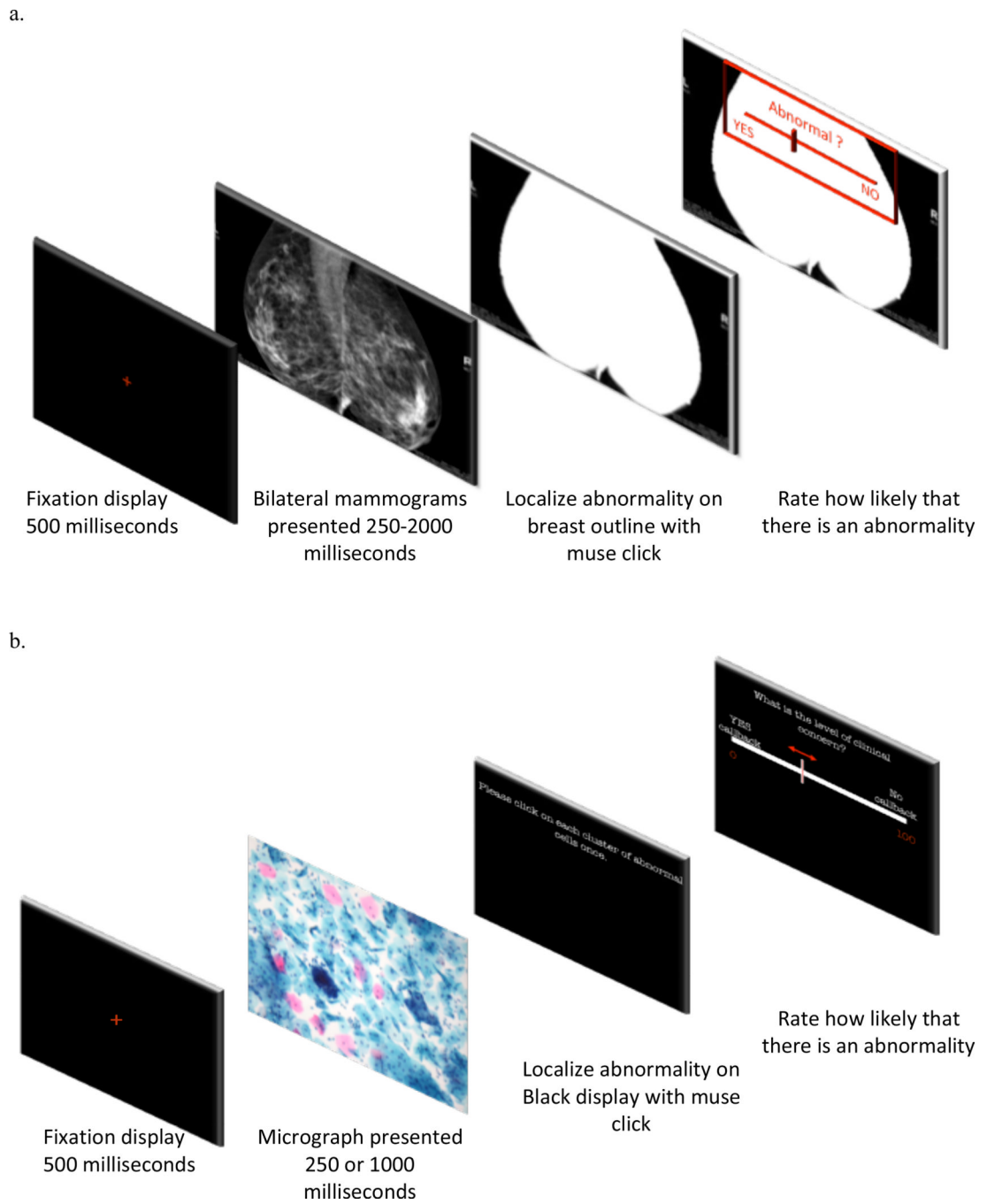
## Acknowledgments

### Funding

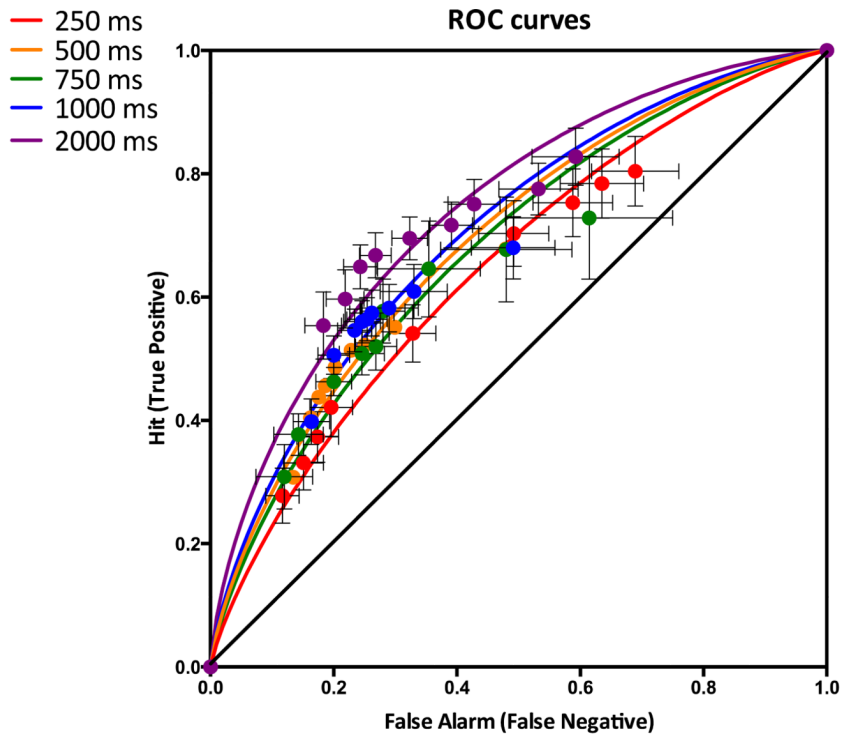
This research was funded by Ruth L. Kirschstein National Research Service Award Grant F32EY019819-01 to Karla K. Evans and by National Institutes of Health–National Eye Institute Grant EY17001 and Toshiba to Jeremy M. Wolfe.

## References

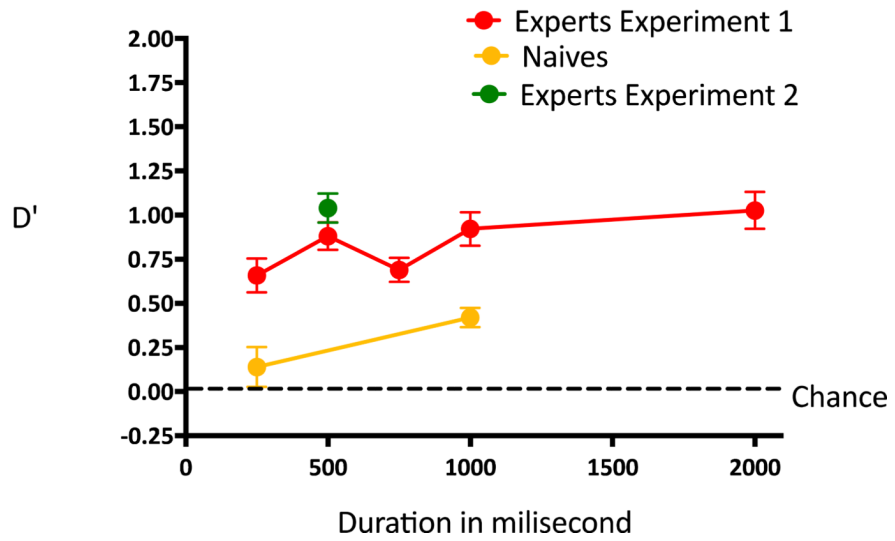
1. Potter MC, Faulconer BA. Time to understand pictures and words. *Nature*. 1975; 253(5491):437–438. [PubMed: 1110787]
2. Intraub H. Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*; *Journal of Experimental Psychology: Human Perception and Performance*. 1981; 7(3):604.
3. Greene MR, Oliva A. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn Psychol*. 2009; 58(2):137–176. [PubMed: 18762289]
4. Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M. Processing scene context: fast categorization and object interference. *Vision Res*. 2007; 47(26):3286–3297. [PubMed: 17967472]
5. Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. *Nature*. 1996; 381(6582):520–522. [PubMed: 8632824]
6. VanRullen R, Thorpe SJ. Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*. 2001; 30(6):655–668. [PubMed: 11464555]
7. Li FF, VanRullen R, Koch C, Perona P. Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci U S A*. 2002; 99(14):9596–9601. [PubMed: 12077298]
8. Evans KK, Horowitz TS, Wolfe JM. When categories collide: accumulation of information about multiple categories in rapid scene perception. *Psychol Sci*. 2011; 22(6):739–746. [PubMed: 21555522]
9. Wolfe JM, Vo ML, Evans KK, Greene MR. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn Sci*. 2011; 15(2):77–84. [PubMed: 21227734]
10. Oliva A, Schyns PG. Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cogn Psychol*. 1997; 34(1):72–107. [PubMed: 9325010]
11. Kundel HL, Nodine CF. Interpreting chest radiographs without visual search. *Radiology*. 1975; 116(3):527–532. [PubMed: 125436]
12. Carmody DP, Nodine CF, Kundel HL. Finding lung nodules with and without comparative visual scanning. *Percept Psychophys*. 1981; 29(6):594–598. [PubMed: 7279589]
13. Oestmann JW, Greene R, Kushner DC, Bourgouin PM, Linetsky L, Llewellyn HJ. Lung lesions: correlation between viewing time and detection. *Radiology*. 1988; 166(2):451–453. [PubMed: 3336720]
14. Mugglestone MD, Gale AG, Cowley HC, Wilson ARM. Diagnostic performance on briefly presented mammographic images. *Medical Imaging*. 1995
15. Mack ML, Palmeri TJ. Modeling categorization of scenes containing consistent versus inconsistent objects. *Journal of Vision*. 2010; 10(3)
16. Rensink RA. Visual sensing without seeing. *Psychological Science*. 2004; 15(1):27–32. [PubMed: 14717828]
17. Simons DJ, Nevarez G, Boot WR. Visual Sensing Is Seeing Why "Mindsight", in Hindsight, Is Blind. *Psychological Science*. 2005; 16(7):520–524. [PubMed: 16008783]
18. Hope, C.; Sterr, A.; Elangovan, P.; Geades, N.; Windridge, D.; Young, K., et al. High throughput screening for mammography using a human-computer interface with rapid serial visual presentation (RSVP); Paper presented at the SPIE Medical Imaging; 2013.



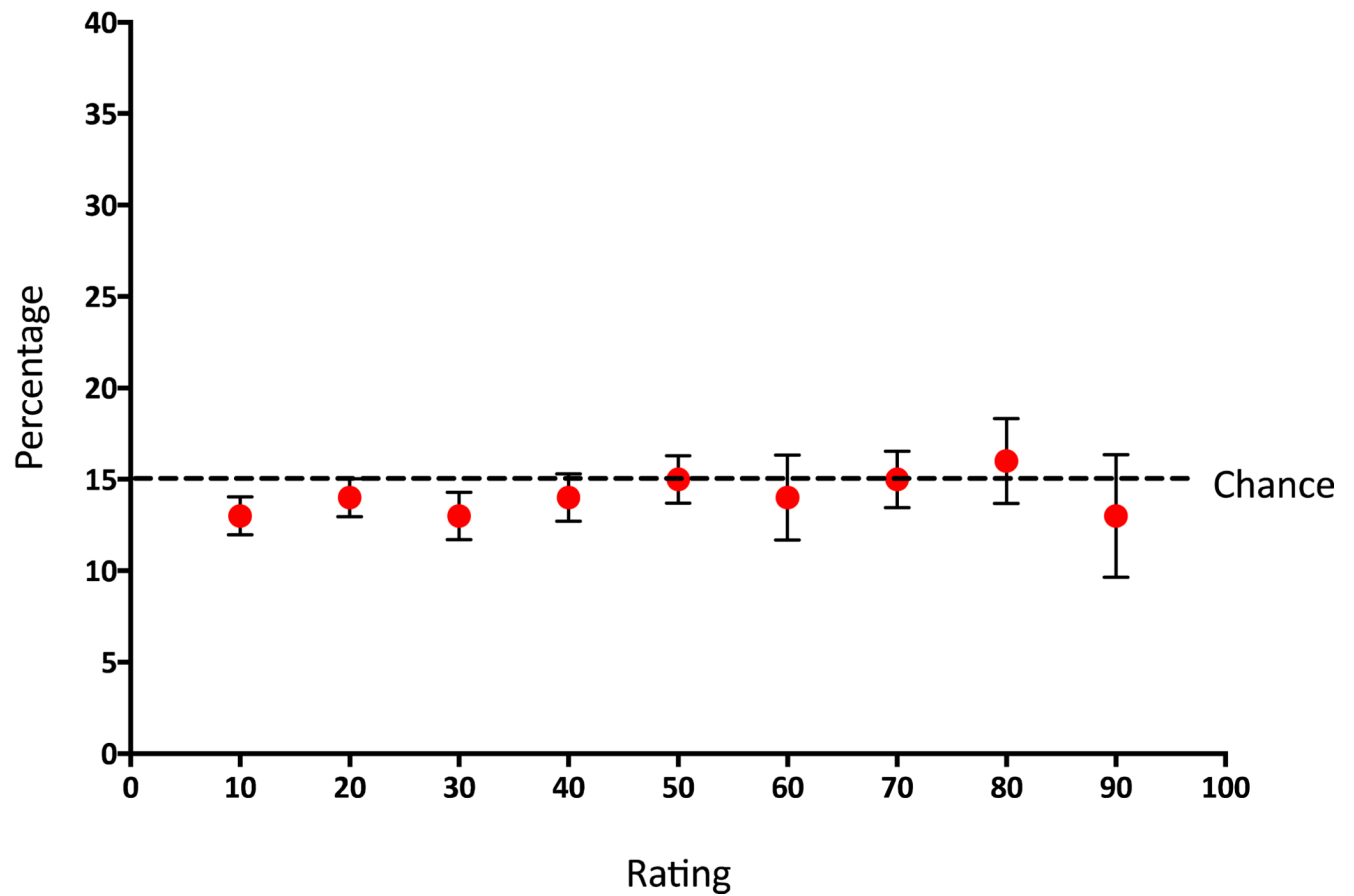
**Figure 1.** Example trials in 3 experiments. a) Example trial presented to radiologists and naïve control group in Experiment 1 & 2. b) Example trial presented to cytologists and naïve control group in Experiment 3.



b.



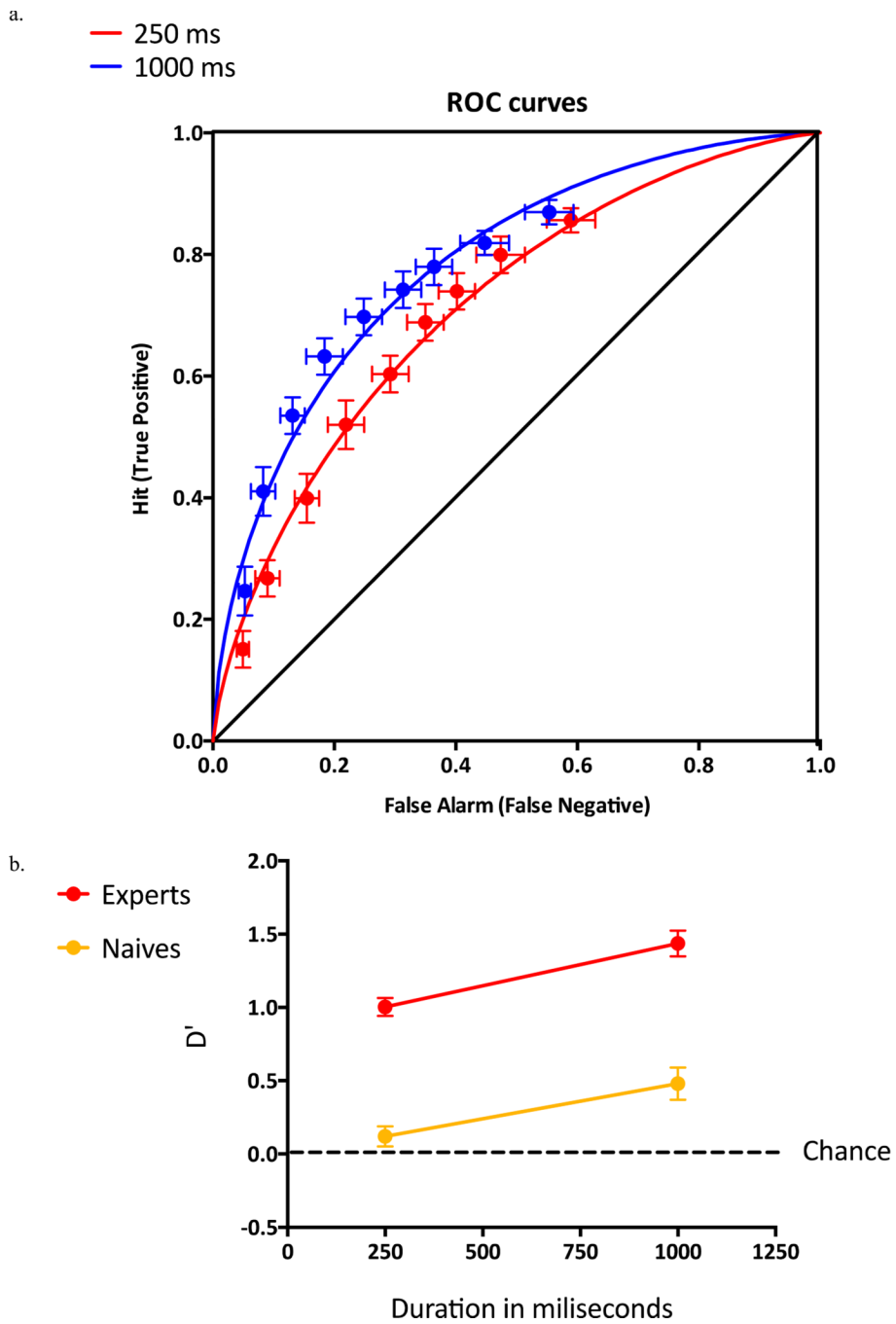
c.



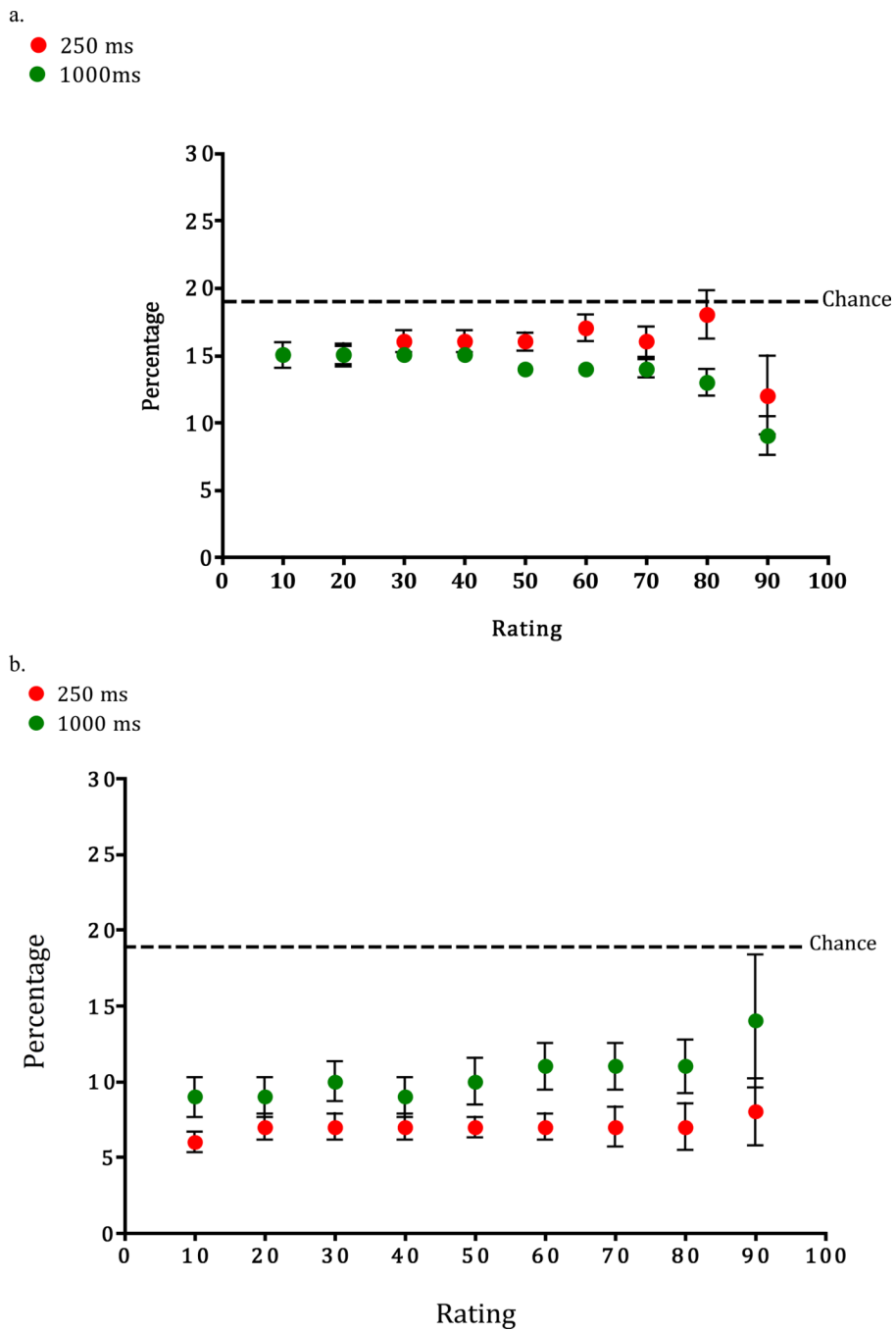
**Figure 2.**

Results of Experiment 1 & 2. a) Rating-based, receiver operating curves (ROC's) of radiologist's performance for each of the five exposure durations in Experiment 1. b) Performance of the expert groups (radiologists in Experiment 1 & 2) and control group (naïve observers) measured in  $D'$  units for each exposure durations. c) Localization performance of radiologists in Experiment 2 across different confidence ratings. Error bars in b and c panel are standard errors of mean.





**Figure 3.** Results of Experiment 3. a) ROCs showing cytologists' performance for each of the two exposure durations. b) Performance of the expert group (cytologists') and control group (naïve observers) measured in  $D'$  units for the two exposure durations. Error bars in panel b are standard errors of mean.



**Figure 4.** Localization performance for a) cytologists and b) naïves in Experiment 3 across different confidence ratings and two exposure durations. Error bars in a and b panel are standard errors of mean.