



Published in final edited form as:

Psychol Sci. 2013 September ; 24(9): 1848–1853. doi:10.1177/0956797613479386.

“The invisible gorilla strikes again: Sustained inattentional blindness in expert observers”

Trafton Drew, Melissa L. H. Vo, and Jeremy M. Wolfe

Trafton Drew: TraftonDrew@gmail.com

Abstract

We like to think that we would notice the occurrence of an unexpected yet salient event in our world. However, we know that people often miss such events if they are engaged in a different task, a phenomenon known as “inattentional blindness.” Still, these demonstrations typically involve naïve observers engaged in an unfamiliar task. What about expert searchers who have spent years honing their ability to detect small abnormalities in specific types of image? We asked 24 radiologists to perform a familiar lung nodule detection task. A gorilla, 48 times larger than the average nodule, was inserted in the last case. 83% of radiologists did not see the gorilla. Eye-tracking revealed that the majority of the those who missed the gorilla looked directly at the location of the gorilla. Even expert searchers, operating in their domain of expertise, are vulnerable to inattentional blindness.

Introduction

When engaged in a demanding task, attention can act like a set of blinders, making it possible for salient stimuli to pass unnoticed right in front of our eyes (Neisser & Becklen, 1975). This phenomenon of “sustained inattentional blindness” is best known from Simons and Chabris’ (1999) study in which observers attend to a ball-passing game while a human in a gorilla suit wanders through the game. Despite having walked through the center of the scene, the gorilla is not reported by a substantial portion of the observers (<http://www.theinvisiblegorilla.com/videos.html>). Does inattentional blindness (IB) still occur when the observers are experts, highly trained on the primary task? There is some evidence that expertise mitigates the effect. For example, Memmert (Memmert, 2006) found a decreased rate of IB for basketball players who were asked to count the number of basketball passes in an artificial game. On the other hand, when Potchen (2006) showed radiologists chest x-rays with a clavicle (collarbone) removed, roughly 60% of radiologists failed to notice when they were reviewing cases as if for an annual exam. Finally, a recent observational case report documented a case where a misplaced femoral line was not detected by variety of health care professional who evaluated the case (Lum, Fairbanks, Pennington, & Zwemer, 2005).

Both of these instances of apparent IB in the medical setting occurred in single-slice medical images. Modern medical imaging technologies like Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Positron Emission Tomography (PET) are increasingly complex: the single image of a chest x-ray has been replaced with hundreds of slices of chest CT scan. It is therefore important to study whether IB occurs in these modern imaging modalities. From the point of view of IB, these situations are interesting because the observer is actively interacting with the stimulus; in this case, scrolling through a stack of

images through the lung. This degree of control may ameliorate the effects of IB because the searcher is able to return and further examine any images that appear unusual.

Moreover, while Potchen showed that radiologists could miss the unexpected absence of a stimulus, we wanted to know if they radiologists would miss a readily detectable, highly anomalous item while performing a task within their realm of expertise. In an homage to the Simons and Chabris' (1999) study, we made that item a gorilla. We compared the performance of radiologists to naïve observers.

Design and Procedure

In computed tomography (CT) lung cancer screening, radiologists search a reconstructed 'stack' of axial slices of the lung for lung nodules that appear as small light circles (Aberle et al., 2011). In Experiment 1, 24 radiologists (mean age: 48; range 28–70), had up to three minutes to freely scroll through each of 5 lung CTs, searching for nodules as we tracked their eyes. Each case contained an average of 10 nodules and the observers were instructed to click nodule locations with the mouse. On the final trial, we inserted a gorilla with a white outline into the lung (see Figure 1A). A typical chest CT 'stack' of images contains 100–500 frames. In the current study, the case that contained the gorilla had 239 slices.

Nine radiologists were tested at Brigham & Women's Hospital (Boston) and 15 were expert examiners from the American Board of Radiology tested at the ABR meeting in Louisville, KY. The gorilla measured 29x50mm. Due to equipment differences, the image size was slightly different at the two sites leading to a small difference in gorilla size (Boston-0.9x0.5 degrees of visual angle, Louisville - 1.3x0.65 DVA). To avoid large onset transients, the gorilla faded into and out of visibility over five, 2mm thick slices of the image (Figure 1). The total volume of rectangular box that could hold the gorilla would be over 7400 mm³, roughly the size of a box of matches. The gorilla was centered near a lung nodule such that both were clearly visible when the gorilla was at maximum opacity. That is, if someone pointed at the correct location in the static image and asked you, "What is that?", you would have no trouble answering, "That is a gorilla". In the scans used in this study, which were taken from the Lung Image Database Consortium (LIDC; Armato et al., 2011), the average volume of lung nodule was 153 mm³. Thus, the gorilla was over 48 times the size of the average nodule in the images (See Figure 2A).

Experiment 2 replicated Experiment 1 with 25 naïve observers (mean age: 33.7; range: 19–55) with no medical training. Prior to the experiment, the experimenter spent roughly 10 minutes teaching the naïve observers how to identify lung nodules. Each experiment began with a practice trial, where the experimenter took time to point out several nodules. They then encouraged the observer to try to find nodules on their own. Once the observer was able to detect at least one nodule, the practice trial was concluded and the experimental trials began. As in Experiment 1, a subset (12) of observers completed the study on a slightly smaller screen. We observed no difference in behavior as result of equipment differences in terms of gorilla or nodule detection.

Experiment 3 was a control experiment to prove that the gorilla was, in fact, visible. Twelve naïve observers (mean age: 37.3; range: 21–54) were shown a movie of the same chest CT case that was used as the final trial in Experiments 1 & 2. The gorilla was inserted on 50% of trials and observers were asked to judge whether the gorilla was present or absent on each of 20 trials. A circular cue indicated the possible location of the gorilla on each trial. The movie played each frame of the case for either 35 or 70ms.

Results

Experiment 1

The nodule detection task was challenging, even for expert radiologists. Overall nodule detection rate was 55%. While engaged in this task, radiologists freely scrolled through the layer containing the gorilla an average of 4.3 times. At the end of the final case, we asked a series of questions to determine whether they noticed the gorilla: “Did the final trial seem any different than any of the other trials?”, “Did you notice anything unusual on the final trial?”, and, finally, “Did you see a gorilla on the final trial?”. Twenty of 24 radiologists failed to report seeing a gorilla. This was not due to the gorilla being difficult to perceive: all 24 radiologists reported seeing the gorilla when asked if they noticed anything unusual on Figure 1 after completion of the experiment (see also exp 3).

The radiologists had ample opportunity to find the gorilla. On average, the radiologists who missed the gorilla spent 5.8 seconds viewing the five slices containing the gorilla (range: 1.1 – 12s). Furthermore, eye-tracking revealed that, of the 20 radiologists who did not report the gorilla, 12 looked directly at the gorilla’s location when it was visible. The mean dwell time on the gorilla amongst this group was 547ms. Figure 2B shows an example from one radiologist who clearly fixated the gorilla but did not report it.

Experiment 2

None of our 24 naïve observers reported noticing the gorilla. As with the radiologist observers in Experiment 1, all of the naïve observers reported seeing the gorilla when shown Figure 1. Similar to Memmert (2006), this pattern of results supports the idea that experts are somewhat less prone to IB (Fisher exact test: $p=0.0497$; see Figure 3A). However, unlike Memmert’s study our two groups showed a sizable difference in performance on the observers’ primary task. As expected, radiologists were much better at detecting lung nodules (mean detection rate: 55%), than naïve observers (12%; $t(47)=12.3$, $p<.001$; see Figure 3B).

Eye-movement data followed the pattern seen with the radiologists. The naïve observers spent an average of 4.9 seconds searching the frames where the gorilla was visible and an average of 157 ms looking in the gorilla’s location. Although both measures show that radiologists who missed the gorilla spent slightly more time searching in the vicinity of the gorilla, neither difference was significant ($t(43)=1.26$, $p=.22$; $t(43)=1.23$, $p=.22$ respectively). Of the 25 naïve observers, 9 looked at the gorilla’s location. The mean dwell time on the gorilla amongst the group that fixated the gorilla was 435ms.

Experiment 3

Although all of our observers in Experiment 1 and 2 reporting seeing the gorilla when shown Figure 1 at end of the experiment, given the very high rate of IB in both studies, there was some concern that the gorilla was too difficult to detect when embedded within a stack of chest CT images. In Experiment 3, each trial consisted of a movie that displayed each level of the chest CT scan from top to bottom. Observers were asked to detect the presence or absence of a gorilla on each trial given a cue to its possible location. Each trial played at a fast or slower frame rate such that the gorilla was visible for 175 or 350ms respectively: substantially less time than the 4.9 seconds that the average naïve observer from Experiment 2 spent searching frames where the gorilla was present. Despite this large difference in time, performance on the detection task was near ceiling (88% correct). Accuracy was not affected by the frame rate ($t(11)=1.1$, $p=.18$, see Figure 3C).

Discussion

In Experiment 1, 20 of 24 expert radiologists failed to note a gorilla, the size of a matchbook, embedded in a stack of CT images of the lungs. This is a clear illustration that radiologists, though they are expert searchers, are not immune to the effects of IB, even when searching medical images within their domain of expertise. Potchen (2006) showed that radiologist could miss the absence of an entire bone. In laboratory search tasks, it is known to be harder to detect the absence of something than to detect its presence (Treisman and Souther, 1985). Our data show that under certain circumstances, experts can also miss the presence of a large, anomalous stimulus. In fact, there is some clinical evidence for errors of this sort in radiology. Lum and colleagues (2005) reported a case study where multiple emergency radiologists failed to detect a misplaced femoral line guidewire that was mistakenly left in a patient and was clearly visible on a chest CT scan. The guidewire was clearly visible on 3 different chest CT scans, but despite being viewed by radiologists, emergency physicians, internists and intensivists, it was not detected and removed for five days. Clearly, radiologists can miss abnormalities that are retrospectively visible when the abnormality is unexpected.

It is reassuring that our experts performed somewhat better than naïve observers as had been reported by Memmert (2006). In that earlier study, expertise was defined as extensive basketball experience and IB was measured during an artificial task where two groups of individuals passed a ball back and forth while moving randomly about a small area. The observers were asked to count the number of passes completed by one group. In this rather abnormal basketball game, the rate of IB was lower for the experts than for those with less basketball experience. In the current study, high rates of IB were obtained with a task and stimulus materials that were very familiar to our expert observers: searching a chest CT scan for signs of lung cancer.

Experts may perform slightly better than naïve observers because their attentional capacity is less completely occupied by the primary task. Simons and Jensen (Simons & Jensen, 2009) recently showed that the rate of IB decreases when the primary task (counting number of object bounces during) is made easier. Along similar lines, there is evidence that training on a specific task reduces subsequent IB rate (Richards, Hannon, & Derakshan, 2010). In our task, the radiologists certainly had much more experience on this specific task, and were clearly better at the task. Both factors are likely to have contributed to the reduced rate of IB observed in our experts. Nevertheless, even though radiologists were slightly better than naïve observers, with an 83% miss rate, the level of IB remains striking.

Why do radiologists sometimes fail to detect such large anomalies? Of course, as is critical in all IB demonstrations, the radiologists were not looking for this unexpected stimulus. In most previous demonstrations of IB, observers engage in a primary task that is unrelated to detection of an unexpected stimulus (such as counting number of passes or bounces, (e.g. Most et al., 2001; Richards et al., 2010; Simons & Chabris, 1999; Simons & Jensen, 2009)). Here, too, though detection of aberrant structures in the lung would be a standard component of the radiologist's task, our observers were not looking for gorillas. Presumably, they would have done much better had they been told to be prepared for such a target. Moreover, the observers were searching for small, light nodules. Previous work with naïve observers shows that IB is modulated by the degree of match between the designated targets and the unexpected item (Most et al., 2001). This suggests that our observers might have fared better if we had used an albino gorilla that better matched the luminance polarity of the designated targets. Counter-intuitively, it could be that a smaller gorilla might have been more frequently detected because it would have more closely matched the size of the lung nodules.

In a radiology context, these results could be seen as an example of a phenomenon known as “satisfaction of search (SoS)”. SoS is a phenomenon in which detection of one stimulus interferes with the detection of subsequent stimuli (e.g. Berbaum et al., 1998). In the present experiment, we placed the gorilla on a slice that contained a nodule that was detected by 71% of our radiologist observers. Perhaps the observed rate of IB was inflated by the presence of this nodule. Without running an additional experiment that examines gorilla detection rate in the absence of the nodule, it is difficult to be certain what role the presence of nodule played. However, if satisfaction of search were truly *driving* the IB effect, we would expect that radiologists who missed the nodule would be more likely to detect the gorilla and that radiologists who found the nodule would be less likely to show IB. Neither of these predictions held true: of the seven radiologists who missed the nodule, none detected the gorilla. Furthermore, all of the radiologists who detected the gorilla also detected the nodule on the same slice.

It would be a mistake to regard these results as an indictment of radiologists. As a group, they are highly skilled practitioners of a very demanding class of visual search tasks. The message of the present results is that even this high level of expertise does not immunize against inherent limitations of human attention and perception. We should seek better understanding of these limits. This would give us a better chance of designing medical and other man-made search tasks in ways that reduce the consequences of these limitations.

References

- Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365(5):395–409. [PubMed: 21714641]
- Armato SG 3rd, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys*. 2011; 38(2):915–931. [PubMed: 21452728]
- Berbaum KS, Franken EA Jr, Dorfman DD, Miller EM, Caldwell RT, Kuehn DM, et al. Role of faulty visual search in the satisfaction of search effect in chest radiography. *Acad Radiol*. 1998; 5(1):9–19. [PubMed: 9442202]
- Lum TE, Fairbanks RJ, Pennington EC, Zwemer FL. Profiles in Patient Safety: Misplaced Femoral Line Guidewire and Multiple Failures to Detect the Foreign Body on Chest Radiography. *Academic Emergency Medicine*. 2005; 12:658–662. [PubMed: 15995100]
- Memmert D. The effects of eye movements, age, expertise on inattention blindness. *Conscious Cogn*. 2006; 15(3):620–627. [PubMed: 16487725]
- Most SB, Simons DJ, Scholl BJ, Jimenez R, Clifford E, Chabris CF. How not to be seen: the contribution of similarity and selective ignoring to sustained inattention blindness. *Psychol Sci*. 2001; 12(1):9–17. [PubMed: 11294235]
- Neisser U, Becklen R. Selective looking: Attending to visually specified events. *Cognitive Psychology*. 1975; 7:480–494.
- Potchen EJ. Measuring observer performance in chest radiology: Some experiences. *Journal of the American College of Radiology*. 2006; 3(6):423–432. [PubMed: 17412097]
- Richards A, Hannon EM, Derakshan N. Predicting and manipulating the incidence of inattention blindness. *Psychol Res*. 2010; 74(6):513–523. [PubMed: 20058156]
- Simons DJ, Chabris CF. Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception*. 1999; 28(9):1059–1074. [PubMed: 10694957]
- Simons DJ, Jensen MS. The effects of individual differences and task difficulty on inattention blindness. *Psychon Bull Rev*. 2009; 16(2):398–403. [PubMed: 19293113]

Gorilla Opacity

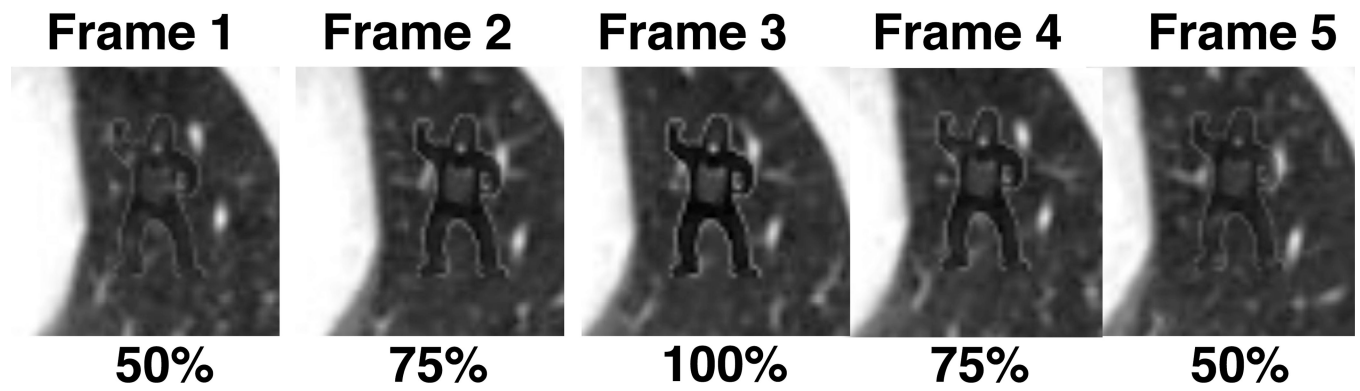


Figure 1.
Gorilla opacity increased from 50 to 100%, then back down to 50% over the course of 5 frames within the chest CT scan.

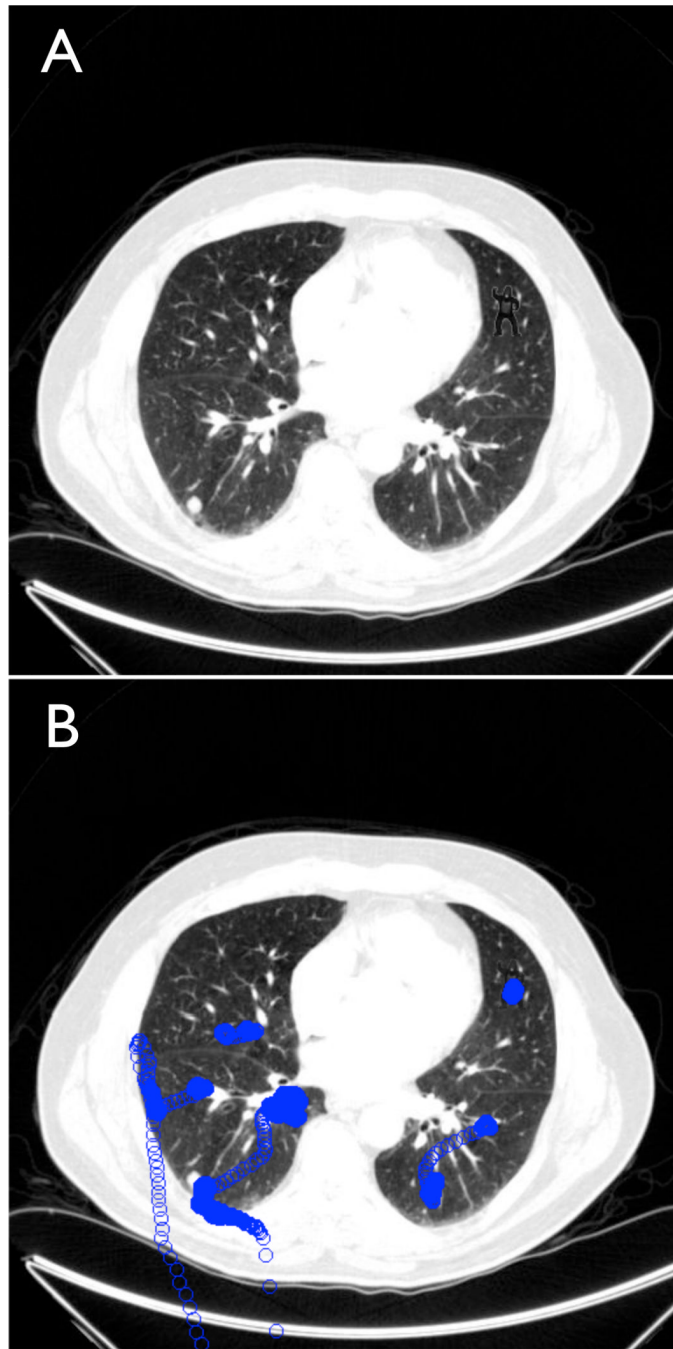


Figure 2.
 A: Chest CT Image containing the embedded gorilla. B: Eye-position plot of one radiologist who did not report seeing the gorilla. Each circle represents eye-position for 1ms.

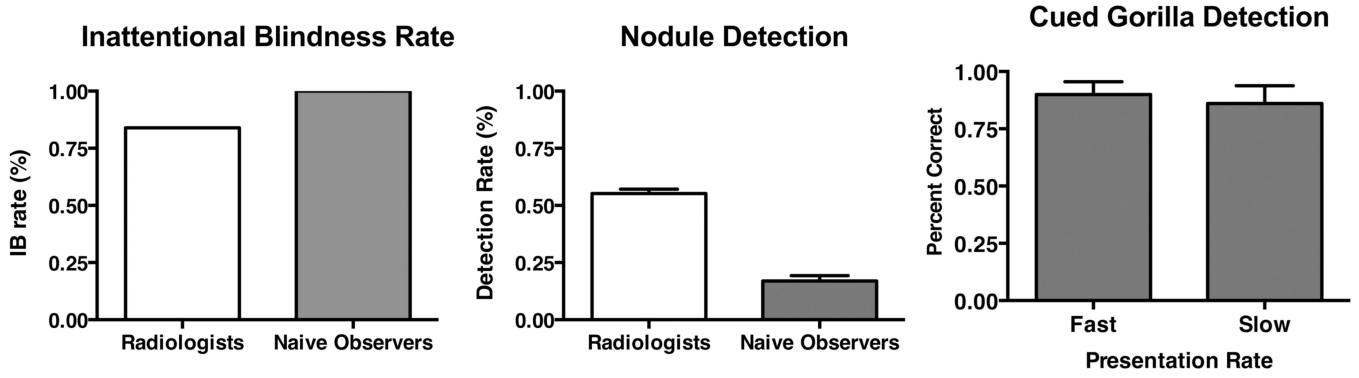


Figure 3. A: Inattentional Blindness rate in Experiments 1 and 2. This represents the percentage of observers who did not report seeing the gorilla. B: Percentage of nodules that were correctly marked by the observers in Experiments 1 and 2. Error bars here and in Figure 3C represent standard error of the mean. C: Gorilla detection rate for Experiment 3 as a function of presentation rate (fast 35 ms/frame or slow 70 ms/frame).