The Invisible Gorilla Strikes Again: Sustained Inattentional Blindness in Expert Observers

Trafton Drew, Melissa L.-H. Võ, and Jeremy M. Wolfe
Visual Attention Lab, Harvard Medical School, and Brigham and Women's Hospital, Boston, Massachusetts

Abstract

Researchers have shown that people often miss the occurrence of an unexpected yet salient event if they are engaged in a different task, a phenomenon known as *inattentional blindness*. However, demonstrations of inattentional blindness have typically involved naive 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 images? We asked 24 radiologists to perform a familiar lung-nodule detection task. A gorilla, 48 times the size of the average nodule, was inserted in the last case that was presented. Eighty-three percent of the radiologists did not see the gorilla. Eye tracking revealed that the majority of those who missed the gorilla looked directly at its location. Thus, even expert searchers, operating in their domain of expertise, are vulnerable to inattentional blindness.

When one is 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 one's eyes (Neisser & Becklen, 1975). This phenomenon of sustained *inattentional blindness* (IB) is best known from Simons and Chabris's (1999) study in which observers attended to a ball-passing game while a human in a gorilla suit wandered through the field of play. Even though the gorilla walked through the center of the scene, a substantial portion of the observers did not report seeing it (the video can be viewed at http://www.theinvisiblegorilla.com/videos.html). Moving beyond such demonstrations, one might ask whether IB still occurs when the observers are experts who are highly trained on the primary task. There is some evidence that expertise mitigates the effect. For example, Memmert (2006) found a decreased rate of IB among basketball players who were asked to count the number of passes in an artificial basketball game. However, when Potchen (2006) asked radiologists to review cases as if for an annual exam and showed them chest x-rays with a clavicle (collarbone) removed, roughly 60% failed to notice the missing bone. Finally, a recent observational report documented that a misplaced femoral line was not detected by a variety of health-care professionals who evaluated the case (Lum, Fairbanks, Pennington, & Zwemer, 2005).

Both of these instances of apparent IB in the medical setting occurred when single-slice medical images were viewed. Modern medical imaging technologies, such as 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 in a chest CT scan. It is therefore important to study whether IB occurs in these modern imaging modalities. These situations are interesting because the observer actively interacts with the stimulus—for example, scrolling through a stack of images of the lung. This degree of control may ameliorate the effects of IB because the searcher is able to return to and further examine any images that appear unusual.

Corresponding Author:
Trafton Drew, Visual Attention Lab, Harvard Medical School, 64 Sidney St., Suite 170, Cambridge, MA 02139
E-mail: tdrew1@partners.org

Moreover, whereas Potchen (2006) showed that radiologists could miss the unexpected absence of a stimulus, we wanted to know if they would miss the presence of a readily detectable, highly anomalous item while performing a task within their realm of expertise. In an homage to Simons and Chabris's (1999) study, we made that item a gorilla. We compared the performance of radiologists with that of naive observers.

Design and Procedure

In CT lung-cancer screening, radiologists search a reconstructed "stack" of axial slices of the lung for 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 3 min to freely scroll through each of five chest CTs, searching for nodules as we tracked their eye position. The five trials contained an average of 10 nodules, and the observers were instructed to click on nodule locations with the computer mouse. In the final trial, we inserted a gorilla with a white outline into the lung (see Fig. 1). A typical stack of images from a chest CT contains 100 to 500 slices. In the current study, the stack that contained the gorilla had 239 slices.

Nine radiologists were tested at Brigham and Women's Hospital in Boston, Massachusetts, and 15 expert examiners from the American Board of Radiology were tested at a meeting of that organization in Louisville, Kentucky. The gorilla measured 29 × 50 mm. Because of equipment differences, the image size was slightly different at the two sites, and consequently the size of the gorilla differed slightly (Boston: 0.9 × 0.5 degrees of visual angle; Louisville: 1.3 × 0.65 degrees of visual angle). To avoid large onset transients, we had the gorilla fade into and out of visibility over five 2-mm-thick slices (Fig. 1). The total volume of the rectangular box that could hold the gorilla was more than 7,400 mm^3, roughly the size of a box of matches. The gorilla was centered in depth 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 (Armato et al., 2011), the average volume of the lung nodules was 153 mm^3. Thus, the gorilla was more than 48 times the size of the average nodule in the images (see Fig. 2a).

Experiment 2 replicated Experiment 1 with 25 naive observers (mean age = 33.7, range = 19–55), who had no medical training. Prior to the experiment, the experimenter spent roughly 10 min teaching these observers how to identify lung nodules. This experiment began with a practice trial, during which the experimenter took time to point out several nodules. The experimenter then encouraged the observer to try to find nodules on his or her 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 gorilla or nodule detection as a result of equipment differences.

Experiment 3 was a control experiment intended to ensure that the gorilla was, in fact, visible. Twelve naive observers (mean age = 37.3, range = 21–54) were shown movies that progressed from the top to the bottom of the same chest CT case that was used as the final trial in Experiments 1 and 2. The gorilla was inserted into the movies in the same location on 50% of the 20 trials, and observers were asked to judge whether the gorilla was present or absent on each trial. A circular cue indicated the possible location of the gorilla on each trial. The movies were presented at a rate of 35 or 70 ms per frame (manipulated within subjects).
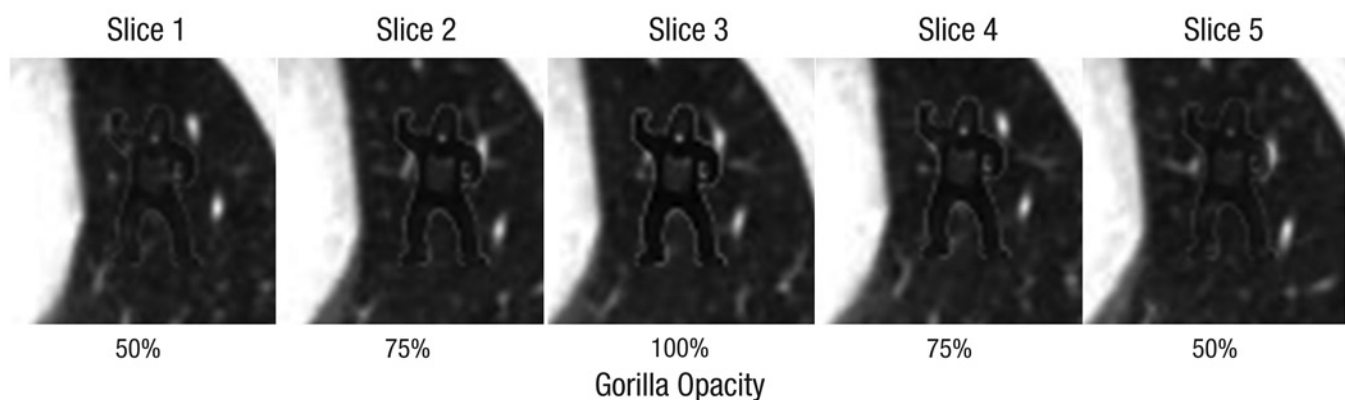
Fig. 1. Illustration of the slices showing the gorilla in the final trial of Experiments 1 and 2. The opacity of the gorilla increased from 50% to 100% and then decreased back down to 50% over the course of 5 slices within a stack of 239.
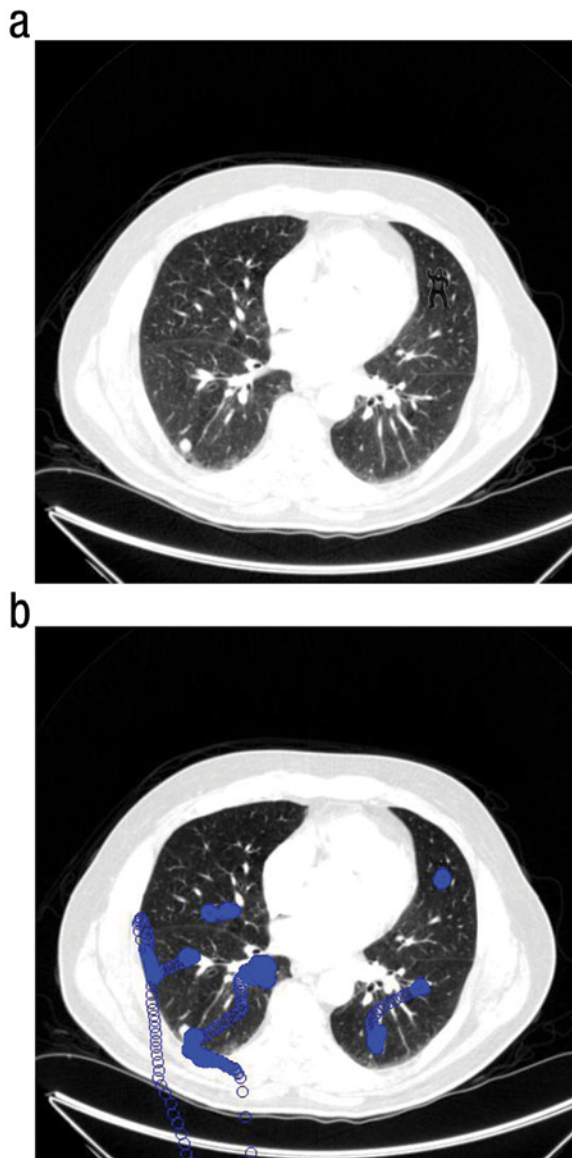
Fig. 2. Computed-tomography image containing the embedded gorilla (a) and eye-position plot of a radiologist who did not report seeing the gorilla (b). In (b), the circles represent eye positions recorded at 1-ms intervals.

Results

Experiment 1

The nodule detection task was challenging, even for expert radiologists. The overall nodule detection rate was 55%. While engaged in this task, the radiologists freely scrolled through the slices 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 had 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 the 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 they were asked if they noticed anything unusual in Figure 1 after completing the experiment (see also the results for Experiment 3).

The radiologists had ample opportunity to find the gorilla. On average, those who missed the gorilla spent 5.8 s viewing the five slices containing it (range = 1.1–12 s) and spent an average of 329 ms looking at the gorilla's location. 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 in this group was 547 ms. Figure 2b shows the eye positions of a radiologist who clearly fixated the gorilla but did not report it.

Experiment 2

None of the 25 naive observers reported noticing the gorilla. As was the case with the radiologists in Experiment 1, all of the naive observers reported seeing the gorilla when shown Figure 1. The results support the idea (Memmert, 2006) that experts are somewhat less prone to IB than novices are (Fisher's exact test: $p = .0497$; see Fig. 3a). However, unlike in Memmert's study, our two groups showed a sizable difference in performance on the primary task. As expected, radiologists were much better at detecting lung nodules (mean detection rate = 55%) than were naive observers (12%), $t(47) = 12.3$, $p < .001$ (see Fig. 3b).

Eye movement data followed the pattern seen with the radiologists. The naive observers spent an average of 4.9 s searching the frames in which the gorilla was visible and an average of 157 ms looking at the gorilla's location. Although both measures showed that radiologists who missed the gorilla spent slightly more time searching in its vicinity than did the naive observers, neither difference was significant, $t(43) = 1.26$, $p = .22$, and $t(43) = 1.23$, $p = .22$, respectively. Of the 25 naive observers, 9 looked at the gorilla's location. The mean dwell time on the gorilla in the latter group was 435 ms.

Experiment 3

Although all observers in Experiments 1 and 2 reported seeing the gorilla when shown Figure 1 at the 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. We tested this possibility in Experiment 3. The movies played at a fast or slower frame rate such that the gorilla was visible for 175 or 350 ms, respectively—substantially less time than the 4.9 s that the average

naive observer in Experiment 2 spent searching frames in which the gorilla was present. Despite this large difference, 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 Fig. 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 radiologists could miss the absence of an entire bone. Results from laboratory search tasks have shown that it is harder to detect the absence of something than to detect its presence (Treisman & 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 et al. (2005) reported a case study in which multiple emergency radiologists failed to detect a misplaced femoral-line guide wire that was mistakenly left in a patient and was clearly visible on three different chest CT scans. Although these scans were viewed by radiologists, emergency physicians, internists, and intensivists, the guide wire was not detected for 5 days. Clearly, radiologists can miss an abnormality that is retrospectively visible when the abnormality is unexpected.

It is reassuring that our experts exhibited somewhat lower rates of IB than naive observers, as was reported by Memmert (2006). In that earlier study, expertise was defined as extensive basketball experience, and IB was measured during an artificial task in which 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
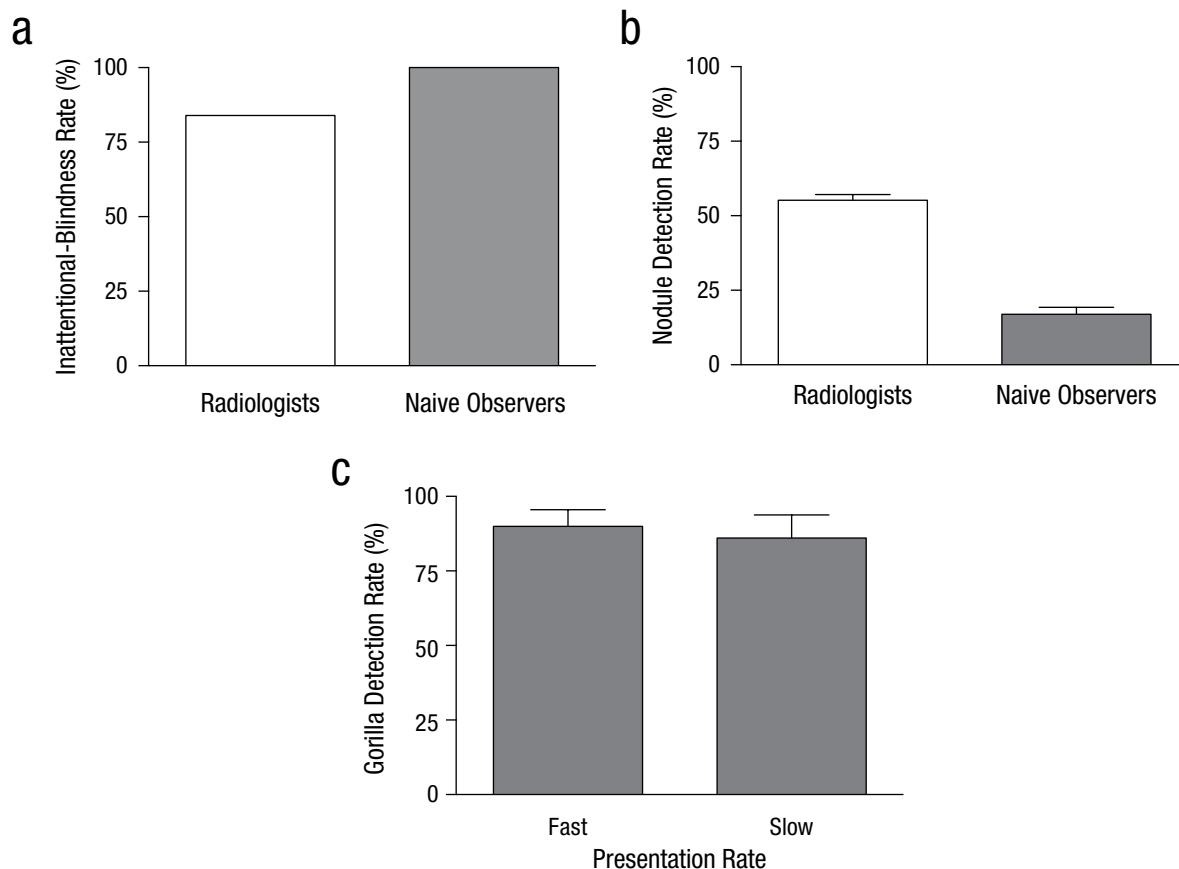
Fig. 3. Experimental results. The graph in (a) shows the rate of inattentional blindness (i.e., the percentage of observers who did not report seeing the gorilla) among the radiologists in Experiment 1 and the naive observers in Experiment 2. The graph in (b) shows the percentage of nodules that were correctly marked by these same observers. The graph in (c) shows the rate at which observers in Experiment 3 detected the gorilla as a function of presentation rate (fast: 35 ms/frame; slow: 70 ms/frame). Error bars represent standard errors of the mean.

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

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 the unexpected stimulus. In most previous demonstrations of IB, observers engaged in a primary task that was unrelated to detection of the unexpected stimulus (e.g., counting the number of passes or bounces, as in 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, observers were not looking for gorillas. Presumably, they would have done much better at detecting the gorilla had they been told to be prepared for such a target. Moreover, the observers were searching for small, light nodules. Previous work with naive observers has shown 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. Counterintuitively, perhaps a smaller gorilla would have been more frequently detected because it would have more closely matched the size of the lung nodules.

Our results could be seen as an example of a phenomenon known as *satisfaction of search*, in which detection of one stimulus interferes with detection of subsequent stimuli (e.g., Berbaum et al., 1998). We placed the gorilla on a slice that contained a nodule that was detected by 71% of the radiologists. Perhaps the observed rate of IB was inflated by the presence of this nodule. Without running an additional experiment examining the detection rate for the gorilla in the absence of the nodule, it is difficult to be certain what role the presence of the nodule played. However, if satisfaction of search truly drove the IB effect, we would expect that radiologists who missed the nodule would have been more likely to detect the gorilla and that radiologists who found the nodule would have been more likely to miss the gorilla. Neither of these predictions held true: Of the 7 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 set of results is that even this high level of expertise does not immunize individuals against inherent limitations of human attention and perception. Researchers should seek better understanding of these limits, so that medical and other man-made search tasks could be designed in ways that reduce the consequences of these limitations.

Author Contributions

T. Drew developed the study concept. All authors contributed to the study design. T. Drew collected and analyzed the data. T. Drew wrote the manuscript in collaboration with J. M. Wolfe and M. L.-H. Võ. All authors approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

References

Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., Fagerstrom, R. M., . . . Sicks, J. D. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, *365*, 395–409.

Armato, S. G., III, McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., . . . Croft, B. Y. (2011). The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, *38*, 915–931.

Berbaum, K. S., Franken, E. A., Dorfman, D. D., Miller, E. M., Caldwell, R. T., Kuehn, D. M., & Berbaum, M. L. (1998). Role of faulty visual search in the satisfaction of search effect in chest radiography. *Academic Radiology*, *5*, 9–19.

Lum, T. E., Fairbanks, R. J., Pennington, E. C., & Zwemer, F. L. (2005). Profiles in patient safety: Misplaced femoral line guidewire and multiple failures to detect the foreign body on chest radiography. *Academic Emergency Medicine*, *12*, 658–662.

Memmert, D. (2006). The effects of eye movements, age, and expertise on inattentional blindness. *Consciousness and Cognition*, *15*, 620–627.

Most, S. B., Simons, D. J., Scholl, B. J., Jimenez, R., Clifford, E., & Chabris, C. F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological Science*, *12*, 9–17.

Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, *7*, 480–494.

Potchen, E. J. (2006). Measuring observer performance in chest radiology: Some experiences. *Journal of the American College of Radiology*, *3*, 423–432.

Richards, A., Hannon, E. M., & Derakshan, N. (2010). Predicting and manipulating the incidence of inattentional blindness. *Psychological Research*, *74*, 513–523.

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

Simons, D. J., & Jensen, M. S. (2009). The effects of individual differences and task difficulty on inattentional blindness. *Psychonomic Bulletin & Review*, *16*, 398–403.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, *114*, 285–310.