*Research Report*

# Rare Targets Are Rarely Missed in Correctable Search

**Mathias S. Fleck and Stephen R. Mitroff**

*Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University*

ABSTRACT—*Failing to find a tumor in an x-ray scan or a gun in an airport baggage screening can have dire consequences, making it fundamentally important to elucidate the mechanisms that hinder performance in such visual searches. Recent laboratory work has indicated that low target prevalence can lead to disturbingly high miss rates in visual search. Here, however, we demonstrate that misses in low-prevalence searches can be readily abated. When targets are rarely present, observers adapt by responding more quickly, and miss rates are high. Critically, though, these misses are often due to response-execution errors, not perceptual or identification errors: Observers know a target was present, but just respond too quickly. When provided an opportunity to correct their last response, observers can catch their mistakes. Thus, low target prevalence may not be a generalizable cause of high miss rates in visual search.*

Whether looking for car keys on a desk or a friend in a crowd, people constantly engage in visual searches of the environment. Ironically, some of the most critical searches often exhibit disturbingly high rates of error: Thirty percent of malignancies are missed in radiological examinations (Berlin, 1994; Renfrew, Franken, Berbaum, Weigelt, & Abu-Yousef, 1992), and a significant percentage of dangerous items are reportedly missed in airport baggage screening. Radiology and airport screening are alike in that the targets of the search are quite rare, and a recent laboratory study (Wolfe, Horowitz, & Kenner, 2005) suggested that low target prevalence per se might directly underlie the high error rates. When searching arrays somewhat similar to those viewed by airport baggage screeners, observers missed only 7% of the targets when target frequency was high (target present on 50% of trials), but an alarming 30% when target frequency was low (target present on 1% of trials).

Address correspondence to Mathias Fleck, Duke University, LSRC Bldg., Box 90999, Durham, NC 27708, e-mail: mathias.fleck@duke.edu.

What drives this potentially dangerous low-prevalence effect? Wolfe et al. (2005) proposed that as observers repeatedly respond with correct rejections (accurately reporting that no target is present on target-absent trials), they begin to terminate their searches more and more quickly, consequently missing targets on the rare trials that actually do contain them. But what is the fate of the target information on those miss trials? Are observers completely unaware of the target, which would suggest that they process misses the same way as correct rejections? Or do observers actually detect the targets, but respond too quickly? In other words, do high miss rates in low-prevalence visual search represent errors of perception or errors of action? This fundamental distinction highlights the importance of the research we report here: We explored the origin of the low-prevalence effect with the direct goal of determining how to eliminate it.

To test whether execution errors account for the increase in misses for rare targets, we used an experimental design similar to that of Wolfe et al. (2005), with the critical modification of providing observers with the opportunity to correct a previous response. Presumably, observers can correct their action-based errors but not their perception-based errors, so with this simple modification, we were able to determine if low target prevalence continues to generate high miss rates when action errors are largely eliminated.

## METHOD

Twenty young adults (average age = 21 years, $SD$ = 4.5 years) were recruited from the Duke University community to participate in the experiment in exchange for $15 or for course credit. All observers gave informed consent prior to participating.

The experiment was conducted on a Dell Optiplex computer running Windows 2000 and programmed in Matlab 6.5 using the Psychophysics Toolbox (Brainard, 1997). Each trial began with a cross ($1.3^\circ \times 1.3^\circ$) appearing for 0.5 s at the center of the screen to indicate the pending onset of the next display. The cross was replaced by the search array, which consisted of 3, 6, 12, or 18 items (see Fig. 1). Items in the search array were 30 photo-realistic objects drawn from the Hemera Photo-Objects

**Fig. 1.** Sample search array. Observers searched for a tool amid randomly rotated items from other categories.

Collections (Hemera Photo Objects, Gatineau, Quebec, Canada). They belonged to five categories: toys, fruits and vegetables, clothing, birds, and tools. Each object was converted to gray scale and partially blurred, then presented with a random rotation in a nonoverlapping array on a white background. The array of possible locations was specified by an invisible $5 \times 5$ grid (subtending $19.1^\circ \times 19.1^\circ$ at an approximate viewing distance of 60 cm), and each item (subtending $3.2^\circ \times 3.2^\circ$, on average) was placed with slight spatial jitter within a randomly selected cell, with the center cell excluded. On target-present trials, one of the items was randomly selected from the tool category (e.g., hammer, wrench, clamp, saw, drill, axe), and the remaining items were drawn randomly, without replacement, from the other four categories. On target-absent trials, all items were drawn from the nontool categories.

Observers searched the display for a tool for as long as they desired, terminating the trial by pressing either the "/" key to indicate target presence or the "Z" key to indicate target absence. Observers were encouraged to treat the experiment as they might an airport security task: It was important to keep the trials progressing, but also imperative that no "dangerous" items (i.e., tools) were missed. Upon response, the display disappeared, and the next trial appeared automatically after a 0.5-s delay. Half of the observers (correction condition) were given the opportunity to correct their response to the previous trial; they were instructed to press the "Esc" key during a trial if their response on the previous trial should be reversed. The other half of the observers had no such option (no-correction condition). Unlike in the study by Wolfe et al. (2005), no feedback was given for a response, nor was there any feedback provided after the correction key was pressed. Observers were told in advance that corrections would be recorded and that after a correction, they should respond to the next trial normally.

The experiment consisted of 1,400 trials divided into three blocks defined by the frequency with which targets were present. The high-prevalence block consisted of 200 trials, 50% of which were target-present trials. The medium-prevalence block also consisted of 200 trials, but in this case a target was present on 10% of trials. The low-prevalence block consisted of 1,000 trials, 2% of which contained a target. Observers were warned that target frequency would generally be very low, and that they should resist any tendency to fall into an automatic "target-absent" response mode. Half of the observers in each condition viewed the blocks in the following order: high prevalence, medium prevalence, low prevalence; the other half viewed the blocks in the reverse order. There was no systematic effect of order in either condition, and all analyses reported here are collapsed over order. After every 200 trials, the program prompted observers to take a break, and the experiment continued when a button was pressed. Each set of 200 trials was preceded by an on-screen indication of the target prevalence (high, medium, or low) in the upcoming set. Observers were strongly encouraged to take advantage of the breaks, particularly if they were feeling tired or bored. The entire experiment ran approximately 80 min in length, depending on the speed of the observer.

## RESULTS AND DISCUSSION

Results for the no-correction condition replicated the prevalence effects of Wolfe et al. (2005). Miss rates were 10%, 19%, and 31% for the high-, medium-, and low-prevalence blocks, respectively, $F(2, 18) = 15.478$, $p < .001$, $\eta_p^2 = .632$ (see Fig. 2). In contrast, the correction condition showed no effect of prevalence, with miss rates of 4%, 10%, and 10% for the high-, medium-, and low-prevalence blocks, $F(2, 18) = 1.618$, $p = .226$, $\eta_p^2 = .152$. A mixed-effects analysis of variance revealed a significant interaction between prevalence and condition, $F(2, 36) = 4.736$, $p = .015$, $\eta_p^2 = .208$, indicating that the prevalence-linked increase in misses occurred specifically when observers could not correct their mistakes. Further, the miss rates in the correction condition calculated *before* incorporating the correction responses (8%, 19%, and 27%, respectively) were statistically equivalent to the miss rates in the no-correction condition: A mixed-effects analysis of variance using these miss rates revealed no interaction between prevalence and condition, $F(2, 36) = 0.186$, $p = .831$, $\eta_p^2 = .010$, again highlighting the specific impact of allowing observers the opportunity to catch their own mistakes.

Average false alarm rates on target-absent trials were very low for all blocks in both conditions; the high-, medium-, and low-prevalence blocks in the no-correction condition yielded false alarm rates of 0.70%, 0.22%, and 0.06%, respectively, and the correction condition produced rates of 0.80%, 0.00%, and

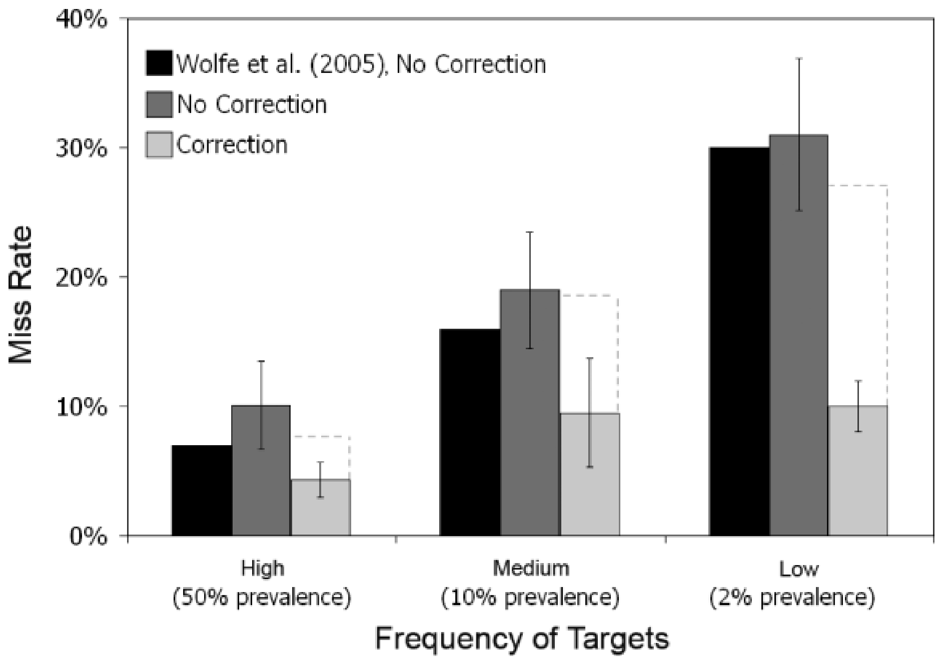


**Fig. 2.** Miss rate as a function of target prevalence. The gray bars show results for the no-correction and correction conditions in the present experiment; results from Wolfe, Horowitz, and Kenner's (2005) study, in which observers did not have the opportunity to correct errors, are reproduced here for comparison. Note that low-prevalence targets appeared on 1% (rather than 2%) of trials in the study by Wolfe et al. The dashed lines indicate the miss rates in the correction condition before observers' corrections were incorporated. Error bars represent ±1 *SEM*.

0.03%, respectively. The correction key was used almost exclusively to correct misses (94.4% of all corrections).

Observers were free to respond at their own pace, and the response time data are highly informative. Figure 3 shows the average response time patterns in the no-correction condition for trials before and after target-present trials in the low-prevalence block, plotted separately for hits and misses. Response times for trials leading up to misses were on average 231 ms faster than response times for trials preceding hits, agreeing with the pattern observed by Wolfe et al. (2005) and supporting the notion that increased speed leads to misses (Chun & Wolfe, 1996; Rabbitt, 1966). The response time data thus seem to show a direct relation between accuracy and speed for this visual search task. Discussing this pattern of results, Wolfe et al. (2005) proposed that misses seem to occur because "observers abandon their search in less than the average time required to find a target" (p. 439). We propose instead that a search may be abandoned in less than the average time required to *respond* to a target, but either way, both sets of data strongly suggest that faster speeds may be responsible for the increase in misses at low target prevalence.

However, the notion of a speed-accuracy trade-off in visual search for rare targets has recently been challenged. New data (Wolfe et al., in press) suggest that when observers are given "speeding tickets" on fast trials to induce slower responding overall, miss rates remain relatively high. If there is a direct link between response time and errors, one would expect improved accuracy in this condition. However, there are a few possible explanations for this discrepancy. First, trial duration in the speeding-ticket experiment was still yoked to response speed, rather than being fixed, and this may have limited any delay-driven benefits by adding a second task. Because observers had to monitor their response speed to avoid penalty, they might have been judging duration while simultaneously trying to complete the search. Second, providing differential feedback for specific durations (punishment after very fast responses and nothing after slower responses) could encourage observers to adopt the strategy of delaying initiation of their search (and thus their response) so as to avoid penalty. Such a process of rescaling a response rule in reference to temporal regularities is similar to mechanisms formalized in information processing models of interval timing (see MacDonald & Meck, 2004, for a review). Finally, it is entirely possible that the induced slowdown simply did not provide enough time to overcome the prepotent "target-absent" response. Although these new results are intriguing, future work will be needed to reconcile them with the current results and an accumulating body of data (e.g., Chun & Wolfe, 1996; Wolfe et al., 2005, in press) that has consistently revealed a relation between faster responses and lower detection rates.

Another interesting pattern is evident in the response time data in Figure 3: Observers in the no-correction condition were on average 161 ms slower to respond on trials immediately

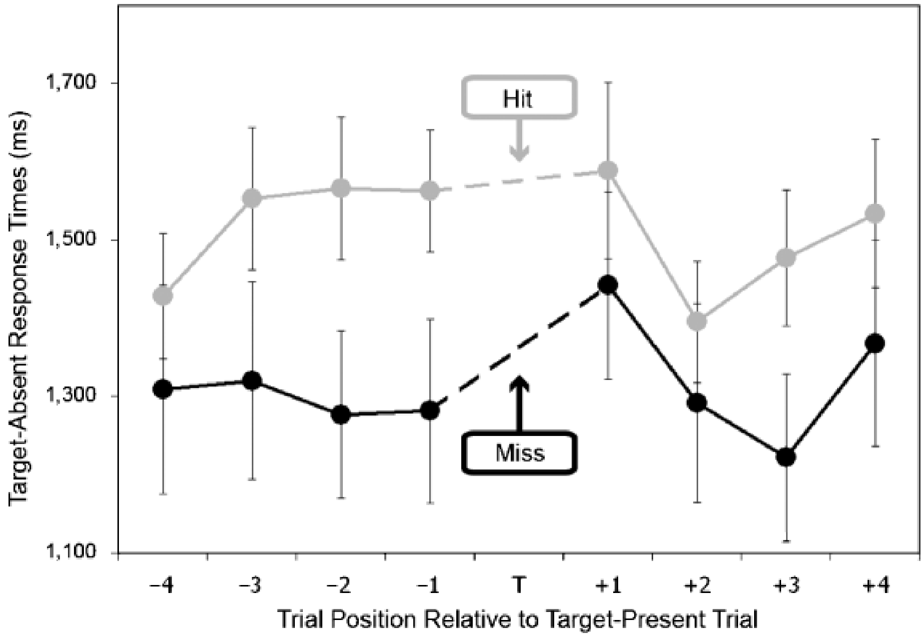


**Fig. 3.** Response time on target-absent trials in the no-correction condition as a function of ordinal relation to a target-present trial. Negative numbers represent successive trials preceding the target (T), and positive numbers represent subsequent trials. Results are shown separately for trials surrounding hits and trials surrounding misses. The dashed lines indicate the change between trials immediately before and after target-present trials. Error bars indicate standard errors of the means.

following a missed rare target than on trials immediately preceding the target—an effect similar to that found by Wolfe et al. (2005) even though in the present case, critically, there was no feedback provided. This slowdown, like the data on the accuracy of corrections in the correction condition, strongly suggests that observers were cognizant of their mistake on some miss trials (Rabbitt, 1966) and that processing may be similar for misses and correct responses (Egeth & Smith, 1967). Thus, these misses appear to be action errors rather than perceptual errors. It should be noted that our miss rate and response time data fully replicate those of Wolfe et al. (2005), which should alleviate concerns about methodological differences (e.g., the presence or absence of feedback).

When given the opportunity, observers readily correct their misses, eliminating the effect of target prevalence in visual search. These findings indicate that the high miss rates in the no-correction condition arose from execution errors; that is, observers in fact noticed these targets but responded too quickly. Whether such late recognition is driven by a lingering sensory representation of the display or reflects an inability to inhibit the repetitive and prepotent "target-absent" response, it is clear that the rise in errors associated with low prevalence is largely driven by a deficit in response execution, rather than by a more general perceptual failure of target identification or search. Because a primary aim of the present work was to relate visual search results to socially important situations, this redefinition of the low-prevalence effect is critical and demands a comparison between the response parameters of laboratory tasks and those of radiological and airport screenings.

In radiology, image readers typically spend 30 to 90 s on an x-ray scan and assess fewer than 100 images in a day, a sharp contrast to the conditions of the present study, which had 1,400 trials and average response times of less than 3 s. Given these differences, misses in this medical context are not likely to be due to the increase in misses caused by rapid responding in tasks with low target frequency, and, indeed, a recent comprehensive radiological study (Gur, Rockette, Armfield, et al., 2003) reported no significant effects of target prevalence on image readers' accuracy in detecting disease. Although target frequency is low in this context, the high incidence of error is likely explained by other mechanisms, including interpretation deficits (Manning, Ethell, & Donovan, 2004), "satisfaction of search" issues (Samuel, Kundel, Nodine, & Toto, 1995; Wolfe et al., 2005), and incomplete visual scan patterns (Kundel, Nodine, & Carmody, 1978).

In contrast to radiological screening, airport baggage screening has relatively fast response times (average inspection times are 3–5 s; Schwaninger, Hardmeier, & Hofer, 2005), and the number of bags screened in a single session can be quite extensive. However, a direct link between baggage screening and our task is tenuous given the differences in response parameters, stimuli, and motivation. Nevertheless, our results underscore the necessity of being able to immediately correct errors (e.g., rewind the baggage conveyor belt or more closely examine individual images) in any fast search for low-prevalence targets. More generally, our results suggest the need to focus less on the effect of prevalence and more on other issues that have been shown to drive high error rates in airport searches, including bag complexity, nonprototypical views of prohibited items, and overlapping x-ray images (Schwaninger et al., 2005), as well as observer-specific factors such as the ability to generalize recognition training to a diverse set of possible threat items (McCarley, Kramer, Wickens, Vidoni, & Boot, 2004).

There remains the possibility that low prevalence may interact with other factors to increase error rates in a manner yet unrevealed. Although our data demonstrate no such interaction, they do highlight the need to minimize or eliminate motor errors when looking for any influence of prevalence. Moreover, establishing that there is no prevalence effect in correctable searches could in fact facilitate the study of misses in searches that may typically involve rare targets (Gur, Rockette, Warfel, Lacomis, & Fuhrman, 2003; Obuchowski, 2005): The large number of trials needed to implement rare-target searches in the laboratory can be extremely cumbersome, and the lack of a prevalence effect indicates that it might be safe to inflate the number of target-present trials to better explore the mechanisms underlying high miss rates.

In sum, prevalence does not influence the error rate in correctable searches. The option to correct mistakes parses out response-execution errors, thus eliminating the rise in miss rates previously found in search for rare targets. Ultimately, improving real-world search performance will be served best by separately addressing errors of action and errors of perception.

*Acknowledgments*—We thank George Alvarez, Greg Appelbaum, Jason Arita, Jessica Cantlon, Ian Dobbins, Chris MacDonald, Greg Stabler, Nick Turk-Browne, Jeremy Wolfe, and two anonymous reviewers for assistance and comments.

## REFERENCES

Berlin, L. (1994). Reporting the "missed" radiologic diagnosis: Medicolegal and ethical considerations. *Radiology*, *192*, 183–187.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 443–446.

Chun, M.M., & Wolfe, J.M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*, 39–78.

Egeth, H., & Smith, E.E. (1967). On the nature of errors in a choice reaction task. *Psychonomic Science*, *8*, 345–346.

Gur, D., Rockette, H.E., Armfield, D.R., Blachar, A., Bogan, J.K., Brancatelli, G., et al. (2003). The prevalence effect in a laboratory environment. *Radiology*, *228*, 10–14.

Gur, D., Rockette, H.E., Warfel, T., Lacomis, J.M., & Fuhrman, C.R. (2003). From the laboratory to the clinic: The "prevalence effect." *Academic Radiology*, *10*, 1324–1326.

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

MacDonald, C.J., & Meck, W.H. (2004). Systems-level integration of interval timing and reaction time. *Neuroscience and Biobehavioral Reviews*, *28*, 747–769.

Manning, D.J., Ethell, S.C., & Donovan, T. (2004). Detection or decision errors? Missed lung cancer from posteroanterior chest radiograph. *British Journal of Radiology*, *77*, 231–235.

McCarley, J.S., Kramer, A.F., Wickens, C.D., Vidoni, E.D., & Boot, W.R. (2004). Visual skills in airport security screening. *Psychological Science*, *15*, 302–306.

Obuchowski, N.A. (2005). One less bias to worry about. *Radiology*, *232*, 302.

Rabbitt, P.M.A. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, *71*, 264–272.

Renfrew, D.L., Franken, E.A., Jr., Berbaum, K.S., Weigelt, F.H., & Abu-Yousef, M.M. (1992). Error in radiology: Classification and lessons in 182 cases presented at a problem case conference. *Radiology*, *183*, 145–150.

Samuel, S., Kundel, H.L., Nodine, C.F., & Toto, L.C. (1995). Mechanisms of satisfaction of search: Eye position recordings in the reading of chest radiographs. *Radiology*, *194*, 895–902.

Schwaninger, A., Hardmeier, D., & Hofer, F. (2005). Aviation security screeners visual abilities & visual knowledge measurement. *IEEE Aerospace and Electronic Systems*, *20*, 29–35.

Wolfe, J.M., Horowitz, T.S., & Kenner, N.M. (2005). Rare items often missed in visual searches. *Nature*, *435*, 439–440.

Wolfe, J.M., Horowitz, T.S., Van Wert, M.J., Kenner, N.M., Place, S.S., & Kibbi, N. (in press). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*.

(RECEIVED 1/8/07; REVISION ACCEPTED 3/2/07; FINAL MATERIALS RECEIVED 3/6/07)