Published in final edited form as:
J Vis. ; 8(15): 15.1–1517. doi:10.1167/8.15.15.

Why do we miss rare targets? Exploring the boundaries of the low prevalence effect

Anina N. Rich,
Macquarie Centre for Cognitive Science, Macquarie University, NSW, Australia, Visual Attention Laboratory, Brigham and Women's Hospital, Boston, MA, USA, & Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Melina A. Kunar,
Department of Psychology, The University of Warwick, Coventry, UK

Michael J. Van Wert,
Visual Attention Laboratory, Brigham and Women's Hospital, Boston, MA, USA

Barbara Hidalgo-Sotelo,
Brain and Cognitive Sciences, MIT, Cambridge, MA, USA

Todd S. Horowitz, and
Visual Attention Laboratory, Brigham and Women's Hospital, Boston, MA, USA, & Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Jeremy M. Wolfe
Visual Attention Laboratory, Brigham and Women's Hospital, Boston, MA, USA, & Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Abstract

Observers tend to miss a disproportionate number of targets in visual search tasks with rare targets. This 'prevalence effect' may have practical significance since many screening tasks (e.g., airport security, medical screening) are low prevalence searches. It may also shed light on the rules used to terminate search when a target is not found. Here, we use perceptually simple stimuli to explore the sources of this effect. Experiment 1 shows a prevalence effect in inefficient spatial configuration search. Experiment 2 demonstrates this effect occurs even in a highly efficient feature search. However, the two prevalence effects differ. In spatial configuration search, misses seem to result from ending the search prematurely, while in feature search, they seem due to response errors. In Experiment 3, a minimum delay before response eliminated the prevalence effect for feature but not spatial configuration search. In Experiment 4, a target was present on each trial in either two (2AFC) or four (4AFC) orientations. With only two response alternatives, low prevalence produced elevated errors. Providing four response alternatives eliminated this effect. Low target prevalence puts searchers under pressure that tends to increase miss errors. We conclude that the specific source of those errors depends on the nature of the search.

Keywords

attention; eye movements; search; target prevalence; feature search; conjunction search

---

Corresponding author: Dr. Anina N. Rich. arich@maccs.mq.edu.au. Address: Macquarie Centre for Cognitive Science, Macquarie University, NSW 2109, Australia.

Commercial relationships: none.

Introduction

Visual search is an important component of everyday life. We look for keys, for cups, for a particular book on the shelf. As a general rule, we tend to look for things that are fairly common (e.g., milk in the fridge). When we look for things that are rare (e.g., caviar in the fridge), we can abandon the search fairly rapidly because the probability of success is low and the stakes are not particularly high. There are some searches, however, that combine low target prevalence with very high stakes. For example, medical screening tasks like mammography or cytopathology screening (‘Pap tests’) are critically important searches for targets that are only rarely present (typically under 1% of cases; Fenton et al., 2007; Gur et al., 2004; Smith & Turnbull, 1997). Similarly, in airport baggage screening, a serious threat in a bag is a highly unusual occurrence (Rubenstein, 2001). Nevertheless, missing these targets can have serious consequences.

Wolfe, Horowitz, and Kenner (2005) demonstrated that when a target is rare, participants are surprisingly poor at detecting it in complex visual displays. Their displays consisted of overlapping, semi-transparent grayscale photographs on a noise background. When the target (a tool) was present on 50% of trials, participants made few ‘miss’ errors (a proportion of 0.07 of target-present trials). In contrast, when the target was only present on 1% of trials, error rates increased dramatically (0.30 of target-present trials). In addition, reaction times (RTs) for target-absent responses were much faster at 1% than at 50% prevalence. At 1%, mean RT for correct target-absent trials was actually less than mean RT for correct target-present trials, a highly unusual pattern for visual search.

This result might suggest a simple speed-accuracy trade-off, but subsequent work points to a more complicated story. The task of Wolfe et al. (2005) produced very few false alarm errors. More recent experiments have used a more difficult task with simulated x-ray images of luggage (Wolfe et al., 2007). In this task, participants made false alarms as well as miss errors, allowing computation of signal detection performance measures. Wolfe et al. (2007) found that the dramatic increase in miss errors at low prevalence was accompanied by a decrease in the false alarm rate. In signal detection terms, this corresponded to a shift in criterion without a significant change in sensitivity (d'), whereas a straightforward speed-accuracy tradeoff would result in a loss of sensitivity as speed increased. This result is consistent with other work in the decision literature showing robust shifts in criterion with changes in the probability of one choice over another (Healy & Kubovy, 1981; Maddox, 2002).

Alternatively, observers might simply make the wrong response. In a low prevalence task of the sort described here, observers will make the same target-absent response on the vast majority of trials. The resulting response priming (Bertelson & Tisseyre, 1966) might cause observers to press the target-absent key when they meant to press the target-present key. Fleck and Mitroff (2007) and Li et al. (submitted) used tasks similar to the original Wolfe et al. (2005) task and found that the excess miss errors at low prevalence could be largely eliminated if observers were forced to slow down or permitted to correct their errors. Such correctable errors might occur either because the observer responded with the wrong motor act or because the observer decided on a target-absent response and then subsequently detected the target too late to withhold the response. In other studies, however, neither slowing participants down (Wolfe et al., 2007) nor permitting correction (Van Wert, Wolfe, & Horowitz, in press) eliminated the prevalence effect in a more difficult baggage screening task.

The diversity of low prevalence results reflects the fact that search termination is a complex process when a target is not found. How do you know when to quit? Miss errors occur when a search task is terminated before the target is found. From a signal detection vantage point, rather than being a single decision, a search task is really a sequence of decisions. For each item, the observer must decide whether or not it is a target. If the decision is ‘target’, then the observer can produce a target-present response. If not, the observer must make a second decision to continue searching or to terminate the search with a target-absent response.

It has proven difficult to produce a fully satisfying model of the decision to terminate an unsuccessful search with an absent response (Chun & Wolfe, 1996; Hong, 2005; Zenger & Fahle, 1997). The data suggest that participants engage in an adaptive process that leads them to terminate the next search sooner after a successful search, and later after an error. It is not clear what is being adjusted in this process. It could be search time (modulated by some estimate of set size), or the proportion of items examined during search. A threshold might be set in terms of some internal measure of salience or activation (i.e., examine all items above some criterion level of interest). It may be that several of these adjustments are made simultaneously. At low prevalence, the input to this adaptive process will be quite extreme. The vast majority of responses will be target-absent responses and most of these would be correct even if the observer did not bother to examine the display. This suggests that any search task will show effects of low prevalence.

In the present paper, we employ much simpler search stimuli than those used in the studies cited above. We show that the effects of prevalence are indeed ubiquitous, elevating miss errors in even the simplest search tasks, and that, as in the work of Fleck and Mitroff and Li et al., the manifestation of the prevalence effect differs depending on the nature of the search. Of course, the simple stimuli used here are far removed from the real-world search tasks that are an important motivation for this line of work. However, these well-controlled stimuli are useful to uncover the fundamental processes that govern search termination on target-absent trials.

In Experiment 1, observers searched for a T among Ls. This spatial configuration search is typically inefficient, with target-present RTs increasing about 20–30 ms for each additional distractor (e.g., Treisman & Gelade, 1980; Wolfe & DiMase, 2003). Nevertheless, once an item is attended, it is trivial to distinguish a T from an L. Such tasks produce very few false alarm errors. In Experiment 2, the task was a feature search for a horizontal line target among vertical line distractors. Here, there were virtually no false alarms and the task does not require much search—in fact, it is effectively a ‘pop-out’ target detection task. To anticipate our findings, we replicated the prevalence effect with these simple, easily discriminable stimuli. Both spatial configuration search (Experiment 1) and feature search (Experiment 2) produced higher miss error rates at low (2%) target prevalence than at high (50%) target prevalence.

Although the first two experiments both showed an effect of target prevalence, the pattern of results differed, suggesting different sources of the effect. Eye movement data in Experiment 1 are consistent with the idea that most of the additional low prevalence errors during spatial configuration search were produced when participants stopped the search early, before fixating the target. Relatively few appeared to be motor errors, where the participants fixated the target but inadvertently pressed the wrong button. In contrast, the additional low-prevalence errors in feature search seemed more likely to be due to motor response errors, since the search component is trivial in this task. This account was tested directly in Experiment 3, where we introduced a minimum duration before participants were allowed to respond. This should primarily reduce motor errors. Consistent with our hypothesis, an enforced delay eliminated the prevalence effect in feature search, but not spatial configuration search.

Finally, in Experiment 4, we show that it is not the simple prevalence of the rare item that determines the rate of miss errors for that target. The target, now present on every trial, was a T which could be upright, rotated 90° to the left or the right of vertical, or inverted. Participants were asked to identify the orientation of the target. When the rare orientation appeared on only 2% of trials, and a single alternative orientation was present on the other 98% of trials, there was a prevalence effect. However, when the rare orientation was presented among three equally common alternative orientations, there was no effect of the prevalence manipulation indicating an important role for the nature of the response, at least when the search task is perceptually simple.

The prevalence effect is not peculiar to perceptually challenging X-ray-like stimuli. Low target prevalence puts pressure on all types of search tasks, changing search behavior in several ways. The relative contribution of these factors differs depending on the structure of the task and the nature of the stimuli.

Experiment 1: Prevalence effects in spatial configuration search

In Fleck and Mitroff (2007) and Wolfe et al. (2005, 2007), the search stimuli were heterogeneous complex objects. In this first experiment, we explore whether the prevalence effect is a general property of visual search, or if it can only be observed when perceptual decisions about each item are relatively difficult. The aim of Experiment 1 was to determine whether the prevalence effect occurs when search is ‘inefficient’, apparently requiring attention to each item in turn, but where the perceptual difficulty of the target/distractor decision is minimized. Participants searched for a T among offset Ls (Figure 1). Once attended, a capital letter T is easily distinguished from an L, even when the lines of the L are slightly offset. Without focal attention, however, the difference in the spatial configuration of the line segments is not available to guide attentional deployment, making search for a T among Ls inefficient (e.g., Treisman & Gelade, 1980; Wolfe & DiMase, 2003). Here, we tested whether this perceptually simple task still produced a prevalence effect when observers had to search for a T that was present on only a small fraction of trials.

We also measured eye position, making it possible to address a number of questions about the pattern of search under different prevalence conditions. First, by measuring number and duration of fixations, we can ascertain if faster response times under low prevalence conditions are due to a reduction in the number of items scanned, or in the time spent on each item. Second, by assessing the spatial distribution of fixations, we can ask if participants are searching the whole display or restricting themselves to a brief search around fixation. Finally, we can ask if, and when, the participants fixated the target item.

Methods

Participants—There were 10 participants (Mean age: 24 years; *SD*: 6.2), all with at least 20/25 visual acuity (with corrective lenses if necessary). All participants gave informed consent and were paid $15/hour.

Apparatus—Displays were presented on a 16″ Trinitron-Dell CRT monitor running at a refresh rate of 100 Hz, controlled by a Dell Optiplex GX270 Intel (R); Pentium (R) 4CPU 2.40 GHz computer. Two-dimensional movements of the right eye were recorded by an ISCAN RK-464 video-based eyetracker sampling at 240 Hz. A chin rest was used to stabilize head position. An initial calibration was performed prior to the beginning of each experimental session to ensure tracking accuracy within 0.5 degrees of visual angle. We used Matlab 5.2 and the Psychophysics Toolbox-2 (Brainard, 1997; Pelli, 1997) to control stimulus presentation and data collection.

Stimuli and procedure—The stimuli were white (42 cd/m-sq) rotated Ts and offset Ls of visual angle 1.0° × 1.0° presented on a mid-gray background (19 cd/m-sq; Figure 1). The stroke width was 0.23 deg. Viewing distance was 75 cm. Stimuli were presented on an imaginary 5 × 5 grid within which each cell was 150 pixels wide × 120 pixels high. Each stimulus was centered within the cell and then randomly jittered ±10 pixels.

Each trial began with a central fixation cross. When the participant had maintained fixation for 500 ms, the search display appeared, and remained until response. Within a block, set sizes of 6 or 12 items were randomly interleaved. A single target T in one of four possible orientations (upright, 90° to the left or the right of vertical, or inverted) was present on either 2% (low prevalence blocks) or 50% (high prevalence blocks) of trials. The remaining stimuli were offset Ls, also presented in the four possible orientations. Participants gave a present/absent judgment using two keys on the keyboard. Participants were informed as to whether targets would be common or rare and were given feedback regarding accuracy after each trial. In the low prevalence condition, it was emphasized that although the targets were rare, it was very important to detect them.

Each participant completed 2200 trials over two sessions. Following initial calibration, each session started with a high prevalence block of 20 practice trials. In Session 1, participants then completed a high prevalence block of 160 trials and a low prevalence block of 1000 trials (with optional breaks every 100 trials). The order of blocks was counterbalanced across participants. In Session 2, participants completed the high prevalence practice block followed by a low prevalence block of 1000 trials (again with optional breaks every 100 trials).

Eye-tracking analyses—Eye movement analyses were performed on smoothed eye position data, averaging the raw data (sampled at 240 Hz) within a moving window of 8 data points (33 ms). The beginning and end positions of saccades were detected using an algorithm implementing an acceleration criterion (Araujo, Kowler, & Pavel, 2001). Specifically, the velocity was calculated for two overlapping 17 ms intervals; the onset of the second interval was 4.17 ms after the onset of the first. The acceleration threshold was set at a velocity change of 6°/s between the two intervals. Saccade onset was defined as the time when acceleration exceeded the threshold, and saccade termination was defined as the time when acceleration dropped below the threshold. Fixations were defined as the periods between successive saccades, with a supplemental threshold criterion of minimum 50 ms fixation duration. Successive saccades that occurred within 50 ms of each other were considered to be a continuous saccade.

Results

Data from one participant had to be excluded due to poor eye-tracking calibration. We removed trials on which the duration of the initial fixation was less than 50 ms and trials on which there were more than 4 instances where the eye-tracker ‘lost’ the eye. This resulted in the exclusion of 0.08 (proportion) of high prevalence and 0.07 of low prevalence trials. We also discarded as outliers trials on which RTs were less than 100 ms or greater than 4000 ms. This resulted in the exclusion of a further 0.04 of high prevalence and 0.001 of low prevalence trials. We used arc-sine transformed accuracy data ($y' = \arcsin(\sqrt{y})$) in all statistical analyses to compensate for unequal variances present in binomial data (Hogg & Craig, 1995). Calculation of mean slope for all experiments was based on the group average of individual participant slopes.

Figure 2a shows the mean error rates for low and high prevalence trials. There were far more miss errors under low than high prevalence conditions. This was confirmed by a repeated-measures ANOVA on arcsine-transformed error data for target-present trials (‘misses’) with the factors of Prevalence (low, high), and Set Size (6, 12). There was a significant effect of prevalence ($F(1,8) = 34.2$, $p < 0.001$, *partial eta*$^2 = 0.81$), but no effect of set size ($F(1,8) < 1$, n.s.) and no interaction ($F(1,8) < 1$, n.s.). False alarm rates were too low for sensible analysis, constituting just 4 of 1285 trials at high prevalence and 12 out of 16656 trials at low prevalence.

Figure 2b shows the mean RT for the two prevalence conditions. When targets were rare, participants were faster to respond on target-absent trials than when targets were frequent. The pattern evident in the figure was confirmed by a repeated-measures ANOVA with the factors of Prevalence (low, high), Target Presence (present, absent) and Set Size (6, 12). Although the three-way interaction did not reach significance ($F(1,8) = 4.6$, $p = 0.064$, *partial eta*$^2 = 0.37$), all two-way interactions were significant (Prevalence and Target Presence, $F(1,8) = 72.5$, $p < 0.01$, *partial eta*$^2 = 0.90$; Prevalence and Set Size, $F(1,8) = 30.0$, $p < 0.001$, *partial eta*$^2 = 0.79$; Target Presence and Set Size, $F(1,8) = 52.4$, $p < 0.001$, *partial eta*$^2 = 0.87$).

Post-hoc pairwise comparisons demonstrated that whereas there was no effect of prevalence on target-present RT ($p > 0.1$), participants were faster under low than high prevalence conditions on target-absent trials ($p < 0.001$). Further, while we observed the typical pattern of slower target-absent than target-present responses at high prevalence ($p < 0.001$), there was no difference between these conditions at low prevalence ($p > 0.1$). There were significant effects of set size on RT for both prevalence conditions and for both target-present and target-absent trials (all $p < 0.01$). Correct target-present slopes of the RT × Set Size functions were 50 ms/item for low prevalence and 81 ms/item for high prevalence. Correct target-absent slopes were 107 ms/item for low prevalence and 168 ms/item for high prevalence.

There are only a small number of trials contributing to the miss RT data in Figure 2b, so these means should be interpreted with caution. With this caveat, we can see that RTs for trials on which participants missed the target are similar to correct target-absent responses under both prevalence conditions. Misses were slower than correct target-present responses for high prevalence but not for low prevalence. This was confirmed by a repeated-measures ANOVA with the factors of Prevalence (low, high), Accuracy (hit, miss) and Set-Size (6, 12). There was a significant interaction between Prevalence and Accuracy ($F(1,8) = 9.5$, $p < 0.02$). This interaction was due to a significant difference between hits and misses for high prevalence ($p < 0.001$) but not for low prevalence ($p > 0.3$). (Note: In this analysis the group mean was substituted for one missing data point because one subject had no errors in the high prevalence, set size 12 condition).

Figures 2c and 2d show mean fixation count and duration data for high and low prevalence conditions for correct (upper) and miss (lower) trials. The number of fixations on correct trials follows the pattern of RT. As in other work, the longer the RT, the more fixations (Hooge & Erkelens, 1998; Motter & Belky, 1998; Shen & Reingold, 1999; Zelinsky, 1996; Zelinsky & Sheinberg, 1997). This pattern was confirmed with a repeated-measures ANOVA with factors of Prevalence (low, high), Target Presence (present, absent) and Set Size (6, 12).

Following a significant three-way interaction ($F(1,8) = 7.0$, $p < 0.05$, *partial eta*$^2 = 0.47$), we performed simple main effects analyses on each prevalence condition separately. For high prevalence data, there was a significant Target Presence by Set Size interaction ($F(1,8) = 44.5$, $p < 0.001$, *partial eta*$^2 = 0.85$). Pairwise comparisons demonstrated that, as one would expect, there were more fixations in set size 12 than set size 6 for both Target Presence conditions ($p < 0.001$). There were also significantly more fixations on target-absent than target-present trials ($p < 0.001$).

In contrast, for low prevalence data, while there was also a significant Target Presence by Set Size interaction ($F(1,8) = 25.0$, $p < 0.01$, *partial eta*$^2 = 0.76$), and an effect of set size for both target-present and target-absent conditions ($p < 0.01$ and $p < 0.001$, respectively), there was no difference in the number of fixations on target-present and target-absent trials ($p > 0.2$). There were no effects of these factors on the duration of search fixations for either prevalence condition (Figure 2d; for Prevalence, Target Presence, and all interaction terms, $F(1,8) < 1$, n.s.; for Set Size, $F(1,8) = 1.7$, $p > 0.2$).

What hints can we get regarding the cause of miss errors by looking at the pattern of fixations? Given that participants were making fewer saccades before giving target-absent responses at low prevalence, they may be systematically failing to examine some portion of the display at low prevalence. The periphery is the most obvious candidate.

In Figure 3, fixations in low and high prevalence conditions are binned by distance from the initial central fixation. The distributions look similar, but a two-tailed Kolmogorov–Smirnoff test revealed a significant difference between the distributions ($p < 0.001$). This difference, however, does not seem to be a systematic neglect of the periphery (or any other area), as one might expect, leaving any theoretical significance unclear.

Fleck and Mitroff (2007) have proposed that the prevalence effect is the consequence of execution errors where participants successfully find the targets under low prevalence conditions, but then make the wrong response. They suggest that these errors are either because participants cannot override the frequent (and therefore pre-potent) target-absent response, or because they only see the target (either directly on the screen or in a lingering representation in iconic or sensory memory) after initiating the incorrect target-absent response. If participants do indeed see the target but push the wrong button, either accidentally due to the pre-potent target-absent response, or due to seeing the target too late (after initiation of an target-absent response), the last eye movement on miss trials should be on, or close to, the target. Although eye movements can be dissociated from the locus of attention, in free-viewing studies like this they give a good indication of the region of space to which attention is deployed (e.g., Hoffman, 1996; Moore & Fallah, 2004; Zelinsky, Rao, Hayhoe, & Ballard, 1997).

To investigate this we sorted the miss trials into three categories (see Figure 4):

1. Trials on which the eyes never landed on or near the target (Figure 4a).
2. Trials on which the eyes landed on the target and then went elsewhere before the end of the trial (Figure 4b).
3. Trials on which the eyes landed on the target as the last fixation prior to response (Figure 4c).

In each case, the 150 × 120 pixel cell of the 5 × 5 stimulus grid was used to define the location of the target. Thus, an eye movement counted as ‘landing on the target’ if there was a fixation within the target cell (denoted by the red square in Figure 4).

Although we cannot simply infer causation from a pattern of fixations, different types of miss errors may well correlate with different patterns of fixation. For example, on trials when an observer commits a miss error because she quit searching before finding the target, it is likely she never fixated that target. Of course, there are alternative explanations for such misses (e.g., the observer found the target without ever fixating it, and then produced a motor error). Nevertheless, if we accept that some accounts are more plausible than others, we can use the eye movement record as an indication of the source of elevated miss errors at low prevalence.

If participants miss more targets at low prevalence because they quit searching earlier, then we should see an increase in the proportion of miss trials where the eyes never reached the target. If participants required more evidence to classify a stimulus as a target at low prevalence (a criterion shift), this might lead to an increase in the proportion of miss errors when the participant fixated the target during the trial, then continued on to fixate other stimuli. Finally, if the misses at low prevalence are due primarily to execution errors (either accidentally pushing the wrong button or finding the target after initiating a target-absent response), we should see an increase in proportion of trials where the target is fixated at the end of the trial, but an incorrect, target-absent response is produced. Errors due to noticing a target in iconic memory after the response has been initiated may also fall in the latter category, as the ‘last fixation’ is measured relative to the response. It is not clear how else we might use fixation patterns to identify this class of errors.

If we look at the breakdown of errors across the different miss categories (Table 1), it is clear that the distribution differs between high and low prevalence. Although the low numbers of trials contributing to some of the categories mean we have to be cautious, a chi-square test using the high prevalence values as predictors showed a significant deviation under low prevalence conditions ($\chi^2(2) = 182$, $p < 0.001$). The final column of Table 1 highlights that the largest change for low prevalence is an increase in the number of miss trials on which the participant did not fixate the target, consistent with termination of search before attention was ever directed to the target. Very few additional miss errors occurred where participants fixated the target and failed to recognize it, providing evidence that the criterion for each individual stimulus decision has not shifted. Finally, there is a small increase in the misses on which the participant fixated the target and then responded target-absent, suggesting some contribution from execution errors.

To further examine the source of the errors in this task, we sorted the target-present RTs for each observer at each set size for low and high prevalence into quartiles[1] (i.e., Quartile 1 has the fastest 25% of RTs, Quartile 4 the slowest 25% of RTs). There are not many trials in each cell, but by summing across observers, we can calculate proportion correct as a function of RT quartile, as shown in Figure 5a.

The critical question for present purposes concerns the source of the extra errors in the low prevalence condition. The data in Figure 5a suggest that, at high prevalence, faster RTs are more accurate. This pattern is less clear at low prevalence, especially for set size 6. What does this suggest about the *additional* miss errors at low prevalence? A prime source of miss errors is the abandonment of search before a target is found. If we subtract the low prevalence accuracy from the high prevalence accuracy shown in Figure 5a, we obtain a plot of the change in accuracy as a function of quartile (Figure 5b). The effect of prevalence on this measure is greater in the earlier quartiles. If miss errors occur when observers incorrectly conclude that search has gone on long enough, these data suggest that observers come to that conclusion more rapidly when they are operating at low prevalence. Note that a similar pattern could occur if participants made motor errors more frequently for targets discovered quickly. However, the eye movement data do not support this hypothesis.

We will briefly consider one more question that can be addressed by these data. Figure 6 presents RTs for correct target-absent trials broken into 100 trial bins. The data demonstrate that the low prevalence effect develops rapidly (within 100 trials for set size 12: target-

---

[1] We thank an anonymous reviewer for this suggestion.

absent RTs faster in low prevalence than high prevalence, $p < 0.05$). Within 100 trials of the start of the second session (on another day), participants were responding more quickly to low prevalence target-absent trials than to high prevalence target-absent trials, despite a practice block of 20 high prevalence trials (set size 6: $p < 0.01$; set size 12: $p < 0.001$).

Discussion

This experiment demonstrates that a large reliable prevalence effect across a group of participants can be obtained with easily discriminable targets. Even slightly offset Ls are easily discriminable from Ts. Nevertheless, under the pressure of low prevalence, participants missed around 40% of the rare targets, and responded more rapidly on target-absent trials than under high prevalence conditions. Thus, the effect is not limited to the perceptually complex stimuli used in previous work.

There was no change in the duration of participants' fixations, and only subtle changes in the area of the display searched under low prevalence. However, we did observe a dramatic decrease in the number of items fixated on low relative to high prevalence trials. The effect of low prevalence on RT was evident within 100 trials, and was quickly reinstated after a one-day break.

The RT patterns suggest that when the target is rare, participants search for a shorter time before concluding the target is absent. At high prevalence, RTs for misses and correct target-absent responses are longer than the mean time required to find a target. At low prevalence, the RTs for target-absent responses and misses are markedly reduced. Similarly, the eye-tracking data suggest that many miss errors at low prevalence were attributable to participants terminating the search without ever fixating the target. Additionally, there was some evidence for errors occurring at the response stage. In the next experiment, we removed the effortful search component from the task to isolate the effect of prevalence on the decision and response stage.

Experiment 2: Prevalence effects in feature search

Experiment 1 demonstrated that the prevalence effect can be observed with perceptually simple stimuli. While most misses did not seem to be simple motor errors, there was at least a small contribution from errors of this sort. In Experiment 2, we look for a prevalence effect with the simplest of 'search' tasks: a 'pop-out' feature detection task where a target immediately attracts attention, minimizing the search component.

Methods

Participants—There were 15 new participants (*Mean age*: 28.3 years; *SD*: 6.2), all with at least 20/25 visual acuity (with corrective lenses if necessary). All participants gave informed consent, and were paid $10/hour.

Apparatus—Displays were presented on a 20″ monitor running at 75 Hz with screen resolution set to 1024 × 768, controlled by a G4 Macintosh computer running Mac OS9. The code was written using Matlab 5.2 and Psychophysics Toolbox-2 (Brainard, 1997; Pelli, 1997).

Stimuli and procedure—Observers searched for a horizontal line target among vertical line distractors. The lines were white (42 cd/m-sq) and presented on a mid-gray background (19 cd/m-sq). Lines subtended 1.4° × 0.4° of visual angle (Figure 7). Viewing distance was approximately 57 cm.

Each trial began with a central fixation cross for 400 ms followed by the search display, which remained visible until response. Within a block, set sizes of 6 and 12 items were randomly interleaved. The target horizontal line was present on either 2% (*low prevalence* blocks) or 50% (*high prevalence* blocks) of trials. Participants gave a present/absent judgment using two keys on the keyboard.

Participants were informed as to whether targets would be common or rare and were given feedback regarding accuracy after each trial. In the low prevalence condition, it was emphasized that although the targets were rare, it was very important to detect them. Each participant completed 2180 trials in a single session, beginning with 20 practice trials at high prevalence, followed by either 160 high prevalence trials and 2000 low prevalence trials (split into two 1000 trial blocks), or the reverse, counterbalanced across subjects. Participants were given frequent opportunities for breaks (between blocks and every 200 trials).

Results

Outlier RTs were defined as for Experiment 1. Less than 1% of trials had to be discarded. Figure 8a shows the mean error rates. Even in this effortless feature task, there were more errors at low prevalence than at high prevalence. This was confirmed by a repeated-measures ANOVA with the factors of Prevalence and Set Size on the arc-sine transformed error rates for target-present trials. This analysis revealed a significant effect of prevalence ($F(1,14) = 16.7$, $p < 0.01$, *partial* $eta^2 = 0.54$), no effect of set size ($F(1,14) < 1$, n.s.), and no interaction ($F(1,14) < 1$, n.s.). There were insufficient errors on target-absent trials for analysis.

An analysis by RT quartile (analogous to the one shown in Figure 5 on Experiment 1 data) revealed that virtually all of the miss errors occur in the fastest RT quartile (70 out of 84 total miss errors). There is a hint of a criterion shift in these data, as there were more false alarms (24 out of 1176 trials) at high prevalence than at low prevalence (2 out of 29366 trials). There are far too many empty cells to calculate d' and criterion, c, for each observer. However, if we treat the entire data set as an omnibus observer, d' is 4.0 at high prevalence and 4.9 at low prevalence. Criterion c shifts from a neutral 0.05 at high prevalence to a conservative 1.37 at low prevalence.

Figure 8b shows mean RT for Experiment 2. Again an effect of prevalence is evident, although the pattern is interestingly different from that of Experiment 1 (cf. Figure 2b). In other low prevalence experiments (both Experiment 1 and previous studies), there was a *speeding up* of target-*absent* responses at low prevalence. In this feature task, the primary effect of low prevalence was a *slowing* of correct target-*present* responses, whereas target-absent responses were unaffected by prevalence. A repeated-measures ANOVA with the factors of Prevalence, Set Size and Target Presence on correct RT revealed a significant interaction between Prevalence and Target Presence ($F(1,14) = 22.4$, $p < 0.001$, *partial* $eta^2 = 0.62$). Simple main effects confirmed that participants were significantly slower to respond on target-present trials in low prevalence relative to high prevalence conditions ($p < 0.01$). There was no difference in RT for target-absent trials.

As with Experiment 1, we need to be cautious in interpreting the miss RT data. The number of trials contributing to the miss RT means shown in Figure 8b is extremely small, particularly for high prevalence. In fact, only eight of the subjects had at least one miss in all the conditions. Although we did not perform a statistical analysis of this due to the small number of trials, at high prevalence the miss RTs fall along the same line as both correct target-present and target-absent RTs. At low prevalence, however, the miss RTs appear faster than all other responses (upright open triangles and lower black line in Figure 8b).

Such a pattern, with fast errors relative to (slower) correct responses is consistent with a speed-accuracy tradeoff. The correct target-present slopes of the RT × Set Size functions were −0.9 ms/item for low prevalence and 0.5 ms/item for high prevalence. Correct target-absent slopes were 2.7 ms/item for low prevalence and 1.6 ms/item for high prevalence.

Discussion

There was no effect of set size on either RT or error rate, consistent with the notion that the target 'popped-out' of the display. Participants were able to detect the presence of the target independent of the number of distractors, as with classic feature search paradigms (see Treisman & Gelade, 1980). Nevertheless, low prevalence significantly increased the rate of missed targets, even though overall error rates were lower than in Experiment 1.

Low prevalence places a strain on the normal processes of search. We hypothesize that the effect of low prevalence manifests in different ways depending on the type of task, particularly the degree of effortful search required. The different pattern of RTs in Experiments 1 and 2 are consistent with this notion. In Experiment 1, observers are searching for something that is very rare. Importantly, they do have to *search* and when the target is rare they sometimes abandon search before finding a perfectly detectable target. The target-present RTs are similar at low and high prevalence. The target-absent trials show a more pronounced effect, with low prevalence RTs being much faster than high prevalence RTs. Experiment 2, by contrast, has only a trivial search component— it is a pop-out feature detection task and, for these purposes, it can be considered to be a simple 2 alternative-forced-choice (2AFC) task with either equal or very unequal probabilities of the two responses. As in other work (Krinchik, 1974; Miller, 1998), RTs for less probable responses are slower than RTs for more probable responses. In both experiments, the *excess* miss errors at low prevalence occur predominantly on trials with shorter RTs. Thus both experiments could be considered to be examples of a speed-accuracy trade-off. In Experiment 2, however, there is also evidence for the effects of response probabilities, with the most common response (target-absent) being overall faster than the less common response (target-present). Thus, we propose that the trade-off in Experiment 1 occurs when observers stop searching too soon while the trade-off in Experiment 2 is the consequence of something more like a motor error, or an error of anticipation, presumably due to influences of response probability. Experiment 3 was designed to test the anticipation hypothesis more directly. Experiment 4 manipulates response probability in more detail.

Experiment 3: Reducing the effect of motor errors

If the prevalence effect is due to a tendency to respond 'the target is absent' automatically at low prevalence (one of the alternatives suggested by Fleck & Mitroff, 2007), then we should be able to eliminate the effect by requiring participants to delay their responses. Here, we introduced a minimum delay before participants were allowed to respond. We compared the effect of this delay on spatial configuration search (as in Experiment 1) and feature search (as in Experiment 2).

Methods

Participants—For the spatial configuration search, there were 9 participants (*Mean age:* 25.9 years; *SD:* 5.4). None of these had participated in Experiment 1. For the feature search, there were 16 participants (*Mean age:* 28.4 years; *SD:* 6.7). Five of these also performed Experiment 2: Two before and three after performing Experiment 3. All participants were screened for visual impairments, and had at least 20/25 visual acuity (with corrective lenses if necessary). They gave informed consent and were paid $10/hour.

Apparatus—The apparatus was the same as for Experiment 2.

Stimuli and procedure—Stimuli for the spatial configuration search were the rotated Ts and offset Ls from Experiment 1. Stimuli for the feature search were the horizontal and vertical lines from Experiment 2. The procedure was similar to that of Experiment 2, with two exceptions. First, we used a single set size of 12 items. Second, we introduced a minimum response duration.

For each version (spatial configuration and feature), there were 1120 trials per participant. Each block was preceded by 20 practice trials at the appropriate prevalence level. For high prevalence blocks, the target was present on 50% of trials and there were 80 trials. In the low prevalence block, the target was present on 2% of trials, and there were 1000 trials. Again, breaks were given between blocks and every 200 trials.

Minimum durations were based on the average RT for the low prevalence correct target-absent responses from the first two experiments. In each case, we took the mean RT for this condition, and added 2 standard deviations to get the minimum duration (rounded to nearest 100 ms). We were aiming for a time-frame in which most successful target-absent responses would be complete. Misses tended to be faster than correct target-present trials under low prevalence. We therefore hypothesized that this duration would prevent those very fast miss errors.

For the inefficient T vs. L search, participants were only allowed to respond after 2000 ms had elapsed. For the efficient feature search, the minimum duration was 500 ms. In each case, the display was presented and no response could be recorded until a tone indicated the end of the minimum duration. Following this time, any response terminated the display. As with the previous experiments, participants were informed when the target would be rare and told to concentrate to ensure they did not miss any targets.

Results

As RT was constrained by the enforced delay, only error data were analyzed for this experiment. Note that the imposed delay eliminated short RTs, but participants were not required to respond immediately on hearing the tone, as they would be in a deadline method (Carrasco & McElree, 2001; McElree & Carrasco, 1999).

Interestingly, participants in the spatial configuration condition tended to make more errors at high prevalence here than in Experiment 1 (cf. Figure 2a). The difference between the two experiments is statistically marginal ($t(16) = 1.95$, $p = 0.07$), and probably reflects random differences between groups of participants. At any rate, the data indicate that simply introducing a delay before response does not reduce the high error rates in the relatively difficult T vs. L search.

Critically, even with the enforced delay, participants in the spatial configuration search condition continued to miss a larger percentage of targets at low than at high prevalence (Figure 9a). This was confirmed by a repeated-measures ANOVA with the factor of Prevalence on the arc-sine transformed error data for target-present trials ($F(1,8) = 15.1$, $p < 0.01$, *partial eta*^{2} = 0.653). Across all observers, there were 7 false alarms out of 434 target-absent trials at high prevalence and 4 out of 10065 target-absent trials at low prevalence.

One participant was removed from the feature search analysis due to responding only *target-absent* in the low prevalence condition. For the feature search (Figure 9b), the delay eliminated the difference between high and low prevalence conditions ($F(1,14) = 2.5$, $p >$

0.1). Across all observers, there were 10 false alarms out of 644 target-absent trials at high prevalence and just one out of 15672 target-absent trials at low prevalence.

Discussion

In Experiment 3, enforcing a delay before response eliminated the prevalence effect for a simple feature search but not for an inefficient T among L search. These results support the hypothesis that the low prevalence effect manifests in different ways, depending on the requirements of the search task, even for simple search tasks. In the highly efficient feature search of Experiment 2, low prevalence seemed to increase the rate of execution or anticipation errors. When the response is forcibly delayed, participants have time to inhibit the pre-potent target-absent response, reducing such motor errors. In the inefficient T among L task of Experiment 1, low prevalence miss errors seemed to occur when the observer abandons the search prematurely. Interestingly, simply forcing observers to spend more time in front of the stimulus in Experiment 3 did not eliminate this manifestation of the prevalence effect. This suggests that the prevalence effect in effortful search is due to more than just motor errors.

This finding is consistent with the results of the ‘speeding ticket’ experiment of Wolfe et al. (2007). In that study, a warning was given to participants when their RTs became too fast. This dramatically slowed RTs but failed to reduce the elevated low prevalence miss rates. Thus, very different searches (simulated baggage screening, T among Ls, and vertical among horizontals) all produce prevalence effects. In the feature search case (vertical among horizontal), the additional miss errors seem to be caused by processes like motor errors due to the unequal probability of each response. These errors can be ‘cured’ by enforcing a delay. The miss errors in the other conditions seem to be a different manifestation of the pressure of low prevalence.

In the final experiment, we used a change in the nature of the response to eliminate the prevalence effect for the spatial configuration search (T among Ls).

Experiment 4: Prevalence effects in identification tasks

The standard prevalence experiment confounds several different forms of rarity. The specific target is rare, *any* target is rare, and the target-present response is rare. In this final experiment, we manipulated the nature of the task in order to partially disentangle these factors.

We again employed the inefficient search for a T among Ls, but rather than using the detection task as in previous experiments, a target T was now present on every trial, and participants had to identify the orientation of that target. Thus, in this task, targets *per se* are no longer rare, but one of the target types can be rare. Note that this identification task encourages participants to keep searching, since there will always be a target. In addition, there were two forms of the task. In one, the T could appear in one of two orientations, requiring a two-alternative forced choice (2AFC) response. In the other, four orientations were possible (4AFC). The 2AFC low prevalence condition, like the preceding studies, encourages a strong motor bias toward one response, while the 4AFC condition removes this bias.

Methods

Participants—There were 8 participants in the 2AFC version (*Mean age:* 35.5 years; *SD:* 12.8), and 12 participants in the 4AFC version (*Mean age:* 28.1 years; *SD:* 7.9). One observer was tested in both conditions. Three had been tested in either Experiment 1 or the spatial configuration version of Experiment 3. All participants were screened for visual impairments, and had at least 20/25 visual acuity (with corrective lenses if necessary). They gave informed consent and were paid $10/hour.

Apparatus—The apparatus was the same as in Experiments 2 and 3.

Stimuli and procedure—Stimuli for the spatial configuration search were the rotated Ts and offset Ls described in Experiment 1. A single T was present on every trial. Participants were asked to identify the orientation of the T as quickly and accurately as possible.

2AFC version: In the high prevalence condition, the T was either upright or inverted with equal probability. Participants pressed one key for upright Ts and another for inverted Ts. In the low prevalence condition, one of these orientations (counterbalanced across participants) occurred on only 2% of trials, with the other orientation present on the remaining 98% of trials.

4AFC version: In the high prevalence condition, each of the four possible orientations appeared with equal probability (25%). Participants used the four arrow keys on the keyboard to indicate the orientation of the T (e.g., right arrow for a T with the stem pointing to the right). In the low prevalence condition, one of these orientations was rare (4% of trials), with the other three equally probable (32%). All other aspects of the procedure were identical to Experiment 2.

Results

Outliers were defined as less than 200 ms or greater than 5000 ms; this resulted in less than 1% of trials being discarded. Figure 10 shows the mean error and RT for both 2AFC and 4AFC versions of the task. The categories of correct and incorrect responses are somewhat different from Experiments 1, 2, and 3, since ‘miss’ and ‘false alarm’ categories do not make sense here. At high prevalence, there are simply correct and incorrect responses. At low prevalence, we can distinguish between correct and incorrect responses for rare and common targets. Since the targets themselves were the same in both prevalence conditions, we have classified the high prevalence errors into errors on targets that matched the rare orientation at low prevalence and those that matched the common orientation. Moreover, in the 4AFC condition, common targets could be misidentified as the rare orientation or one of the other two common orientations. There were, however, very few of these errors.

Looking first at the 2AFC error data (Figure 10a), it is clear that there were fewer errors in this version of a T among Ls search than in the versions used in Experiments 1 and 3 (cf. Figures 2 and 9, respectively). The pattern looks more like the pattern of errors seen in the feature search of Experiment 2 (cf. Figure 8a). A repeated-measures ANOVA with the factors of Prevalence and Set Size on the arc-sine transformed misses of the rare target confirmed a significant effect of Prevalence ($F(1,7) = 28.12$, $p < 0.01$, *partial* $eta^2 = 0.80$), but no effect of Set Size ($F(1,7) < 1$, n.s.), and no interaction ($F(1,7) < 1$, n.s.).

Analogous to the elevated low prevalence miss rates in previous experiments, participants were more likely to miss the rare target orientation, incorrectly responding with the frequent response key. There were too few errors on the common targets (analogous to false alarms in the detection experiments) for sensible analysis: In the 2AFC version, across observers, there were only 5 common target errors at high prevalence out of a total of 635 common target trials. At low prevalence, there were 15 common target errors out of 15647 common target trials. In the 4AFC version, across observers, there were only 19 common target errors at high prevalence out of the 1406 common target trials. At low prevalence, there were 92 out of 11320 common target trials.

The correct RTs for the 2AFC (Figure 10c) show a reliable pattern. A repeated-measures ANOVA with the factors of Prevalence, Set Size, and Target Frequency revealed an interaction between Prevalence and Target Frequency ($F(1,7) = 16.13$, $p < .01$, *partial eta*^{2} = 0.70), and a significant effect of Set Size ($F(1,7) = 171.67$, $p < 0.001$, *partial eta*^{2} = 0.96). Under low prevalence conditions, participants were slower to correctly respond to the rare target than the common target ($p < 0.01$), but under high prevalence, there was no difference between the orientations ($p > 0.4$). The correct rare-target slopes of the RT × Set Size functions were 75 ms/item for low prevalence and 56 ms/item for high prevalence. Correct target-absent slopes were 54 ms/item for low prevalence and 59 ms/item for high prevalence.

The 4AFC data were quite different (Figures 10b and 10d). A repeated-measures ANOVA with the factors of Prevalence and Set Size on the arcsine-transformed rare target error (miss) data revealed no significant effects (all $p > 0.15$). In the RT data, with the factors of Prevalence, Set Size and Target Frequency, there was a significant interaction between Prevalence and Target Frequency ($F(1,11) = 5.25$, $p < 0.05$, *partial eta*^{2} = 0.32) as well as a significant effect of Set Size ($F(1,11) = 121.14$, $p < 0.001$, *partial eta*^{2} = 0.92). The interaction appears to be due to RTs for the common target at low prevalence being slightly faster than RTs to the matched orientation at high prevalence, but this effect is weak and was not detectable in the simple main effects ($p > 0.50$). The rare-target slopes of the RT × Set Size functions were 88 ms/item for low prevalence and 81 ms/item for high prevalence. Target-absent slopes were 105 ms/item for low prevalence and 106 ms/item for high prevalence.

Discussion

Like Experiments 1 and 3, the task in Experiment 4 was a search for a T among Ls. In the low prevalence conditions of the earlier experiments, the target stimulus was rare and often missed. We argued that many of those rare target miss errors were caused by early termination of the search, prior to finding the T. In Experiment 4, there was a target on every trial. Thus, participants were encouraged to continue searching until they found a T. As expected, the rare target error rate dropped dramatically under these conditions. This could be due to participants simply guessing that the common target is present, without searching. If this is the case, these correct common target responses should be very fast. On other trials, they may only search for a set time and, if they fail to find the target, guess that the target is the common orientation. This should result in a slower RT for these trials, but still involves guessing. Under either strategy (or a mix of the two), however, the participants would only get the *rare* target correct if they find it before guessing, and therefore the mean RT for correct rare target trials should be faster than the mean RT for the common targets. In contrast, Figure 10c shows the opposite pattern. Thus, the remaining rare target miss errors in the 2AFC case seem to be response errors as seen in the feature search tasks of Experiments 2 and 3. One imagines that participants knew that many of these responses were errors and would have corrected them given the opportunity as in Fleck and Mitroff's (2007) experiments. In the 4AFC case, where the strong bias toward one response is removed, the low prevalence errors vanish, a pattern seen in some studies of response probability (Bertelson & Tisseyre, 1966).

General discussion

This set of experiments demonstrates the ubiquity of low prevalence effects in visual search. The effects can be seen in the simplest search tasks. This should not be too surprising since prevalence effects (under various names) appear in numerous contexts from studies of vigilance (Colquhoun & Baddeley, 1967) to predation by blue jays (Bond & Kamil, 2002).

Under low target prevalence conditions, observers quite reasonably conclude that targets are unlikely to occur. In visual search tasks, one can imagine at least three different responses to that conclusion. First, the observer can prepare to emit a target-absent response. Under conditions where the observer is being asked to respond quickly, that absent response might be given inadvertently from time to time, producing miss errors (Experiments 2 and 4: 2AFC) that are likely to be correctable (Fleck & Mitroff, 2007; it is interesting that the pattern of errors in Fleck & Mitroff's paper suggested response errors in a task that is undoubtedly not a simple feature search). If there are multiple responses, no single motor response will be primed in this way and we do not see the same elevation in miss errors (Experiment 4: 4AFC). Second, if the task requires a search, an observer might abandon the search too soon. Errors of this sort were reflected in the eye movement data of Experiment 1. These errors seem to persist even when observers were required to wait before responding (Experiment 3). Finally, if the discrimination of target and distractor is difficult, the observer would be more likely to conclude that an ambiguous item was a distractor than a target if targets are rare. This would bias observers toward absent responses and miss errors in difficult tasks and would be manifest as a criterion shift. These errors are not seen in the present experiments because the search tasks all involve simple stimuli. Perhaps there is a hint of such an effect in the false alarm data where, in several experiments, more false alarms were found at high than at low prevalence. The small number of false alarms produced by these easy searches makes this result, at best, suggestive. The simulated baggage search task of Wolfe et al. (2007), however, was much more difficult than the tasks used here and produced just this sort of criterion shift.

All of these responses to low prevalence can be seen as sensible and adaptive unless miss errors are much less desirable than false alarms, as they would be in most low prevalence medical screening or airport security settings. Under those circumstances, one might hope that the observer's understanding of the costs of different types of error would counteract the forces of prevalence, acting like a payoff matrix. Pilot studies in our lab did not succeed in eliminating the prevalence effect by explicitly manipulating the monetary payoffs for different types of response. However, Navalpakkam, Koch, and Perona (2007) have had more success. The matter is worth further study since the results of the current experiments make it clear that the pressure of low prevalence will alter behavior in even the most basic of search tasks.

Acknowledgments

This research was supported by grants from the Department of Homeland Security's Transportation Security Laboratory Human Factors Program (USA) and the Air Force Office of Scientific Research. ANR was supported by the National Health and Medical Research Council and the Menzies Foundation (Australia; 359244). BHS was supported by the National Science Foundation (USA). We thank MS Fleck and SR Mitroff for useful comments and A Oliva for access to the eyetracker.

References

Araujo C, Kowler E, Pavel M. Eye movements during visual search: The costs of choosing the optimal path. Vision Research. 2001; 41:3613–3625. [PubMed: 11718799]

Bertelson P, Tisseyre F. Choice reaction time as a function of stimulus versus response relative frequency of occurrence. Nature. 1966; 212:1069–1070. [PubMed: 21090504]

Bond AB, Kamil AC. Visual predators select for crypticity and polymorphism in virtual prey. Nature. 2002; 415:609–613. [PubMed: 11832937]

Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:443–446. [PubMed: 9176954]

Carrasco M, McElree B. Covert attention accelerates the rate of visual information processing. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:5363–5367. [PubMed: 11309485]

Chun MM, Wolfe JM. Just say no: How are visual searches terminated when there is no target present? Cognitive Psychology. 1996; 30:39–78. [PubMed: 8635311]

Colquhoun WP, Baddeley AD. Influence of signal probability during pretraining on vigilance decrement. Journal of Experimental Psychology. 1967; 73:153–155. [PubMed: 6031659]

Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. New England Journal of Medicine. 2007; 356:1399–1409. [PubMed: 17409321]

Fleck MS, Mitroff SR. Rare targets are rarely missed in correctable search. Psychological Science. 2007; 18:943–947. [PubMed: 17958706]

Gur D, Sumkin JH, Rockette HE, Gannot M, Hakim C, Hardesty L, et al. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. Journal of the National Cancer Institute. 2004; 96:185–190. [PubMed: 14759985]

Healy AF, Kubovy M. Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. Journal of Experimental Psychology: Human Perception and Performance. 1981; 7:344–354.

Hoffman, JE. Visual attention and eye movements. In: Pashler, H., editor. Attention. London: University College London Press; 1996.

Hogg, RV.; Craig, AT. Introduction to mathematical statistics. 5. Upper Saddle River, NJ: Prentice-Hall; 1995.

Hong SK. Human stopping strategies in multiple-target search. International Journal of Industrial Ergonomics. 2005; 35:1–12.

Hooge IT, Erkelens CJ. Adjustment of fixation duration in visual search. Vision Research. 1998; 38:1295–1302. [PubMed: 9666997]

Krinchik E. Probability effects in choice reaction time tasks. Perception & Psychophysics. 1974; 15:131–144.

Li, H.; Li, F.; Gao, HH.; Chen, A.; Lin, C. Unpublished manuscript. 2006. Appropriate responding can reduce miss errors in visual search.

Maddox WT. Toward a unified theory of decision criterion learning in perceptual categorization. Journal of Experimental Analysis of Behavior. 2002; 78:567–595.

McElree B, Carrasco M. The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. Journal of Experimental Psychology: Human Perception and Performance. 1999; 25:1517–1539. [PubMed: 10641310]

Miller J. Effects of stimulus-response probability on choice reaction time: Evidence from the laterlized readiness potential. Journal of Experimental Psychology: Human Perception and Performance. 1998; 24:1521–1534.

Moore T, Fallah M. Microstimulation of the frontal eye field and its effects on covert spatial attention. Journal of Neurophysiology. 2004; 91:152–162. [PubMed: 13679398]

Motter BC, Belky EJ. The guidance of eye movements during active visual search. Vision Research. 1998; 38:1805–1815. [PubMed: 9797959]

Navalpakkam, V.; Koch, C.; Perona, P. Optimality of visual search behavior. Paper presented at the Joint Symposium of Neural Computation; 2007.

Pelli DG. The Video Tool box software for visual psychophysics: Transforming numbers into movies. Spatial Vision. 1997; 10:437–442. [PubMed: 9176953]

Rubenstein, J., editor. Test and evaluation plan: X-ray image screener selection test. Washington, DC: Office of Aviation Research; 2001.

Shen, J.; Reingold, EM. Saccadic selectivity during visual search: The effects of shape and stimulus familiarity. In: Hahn, M.; Stoness, SC., editors. Proceedings of the Twenty First Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum; 1999.

Smith PA, Turnbull LS. Small cell and 'pale' dyskaryosis. Cytopathology. 1997; 8:3–8. [PubMed: 9068949]

Treisman AM, Gelade G. A feature-integration theory of attention. Cognitive Psychology. 1980; 12:97–136. [PubMed: 7351125]

Van Wert MJ, Wolfe JM, Horowitz TS. Even in correctable search, some types of rare targets are frequently missed. Perception & Psychophysics. (in press).

Wolfe JM, DiMase JS. Do intersections serve as basic features in visual search? Perception. 2003; 32:645–656. [PubMed: 12892426]

Wolfe JM, Horowitz TS, Kenner NM. Cognitive psychology: Rare items often missed in visual searches. Nature. 2005; 435:439–440. [PubMed: 15917795]

Wolfe JM, Horowitz TS, Van Wert MJ, Kenner NM, Place SS, Kibbi N. Low target prevalence is a stubborn source of errors in visual search tasks. Journal of Experimental Psychology: General. 2007; 136:623–638. [PubMed: 17999575]

Zelinsky GJ, Rao RPN, Hayhoe MM, Ballard DH. Eye movements reveal the spatiotemporal dynamics of visual search. Psychological Science. 1997; 8:448–453.

Zelinsky GJ. Using eye saccades to assess the selectivity of search movements. Vision Research. 1996; 36:2177–2187. [PubMed: 8776484]

Zelinsky GJ, Sheinberg DL. Eye movements during parallel-serial visual search. Journal of Experimental Psychology: Human Perception and Performance. 1997; 23:244–262. [PubMed: 9090154]

Zenger B, Fahle M. Missed targets are more frequent than false alarms: A model for error rates in visual search. Journal of Experimental Psychology: Human Perception and Performance. 1997; 23:1783–1791. [PubMed: 9425680]

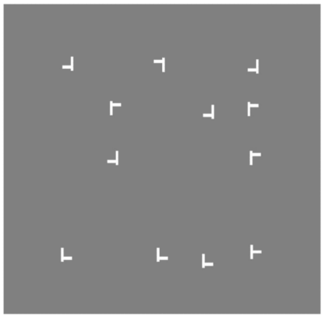

Figure 1.

Sample target-present display for Experiment 1. Participants searched for a T in any orientation among offset L distractors, forming a difficult spatial configuration search.

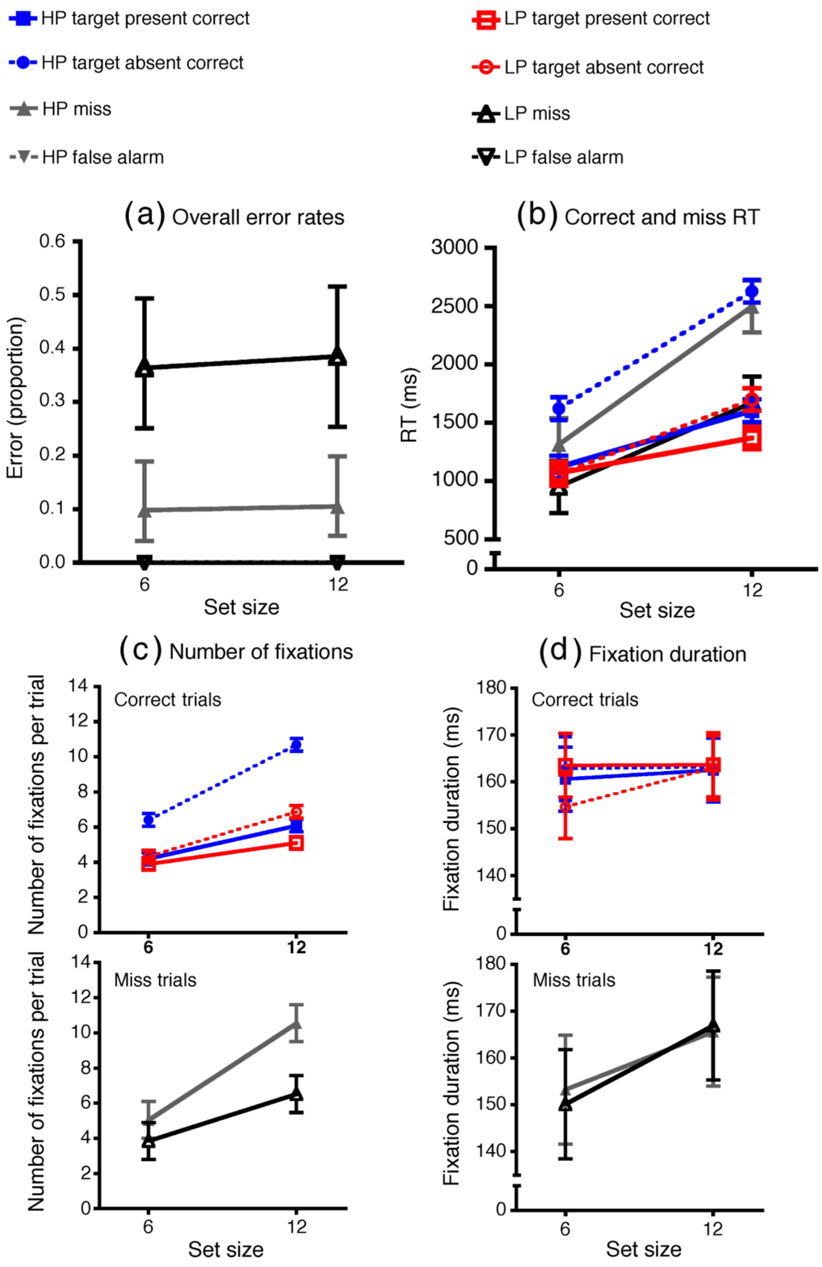

Figure 2.

Results from Experiment 1. (a) Mean error rate (proportion), (b) mean RT (ms), (c) mean number of fixations, and (d) mean fixation duration (ms), for correct and incorrect responses for each of the prevalence and target presence conditions, plotted by set size. Error bars represent within-subjects 95% confidence intervals. Note the fixation data are from search fixations only (excludes the initial central fixation for all conditions and the target fixation for target-present trials). HP = high (50%) prevalence, LP = low (2%) prevalence.

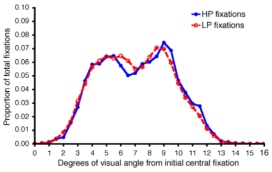

Figure 3.

Spatial distribution of fixations relative to the initial central fixation, plotted separately for high (blue) and low (red) prevalence, and collapsed across the other conditions. HP = high (50%) prevalence, LP = low (2%) prevalence.

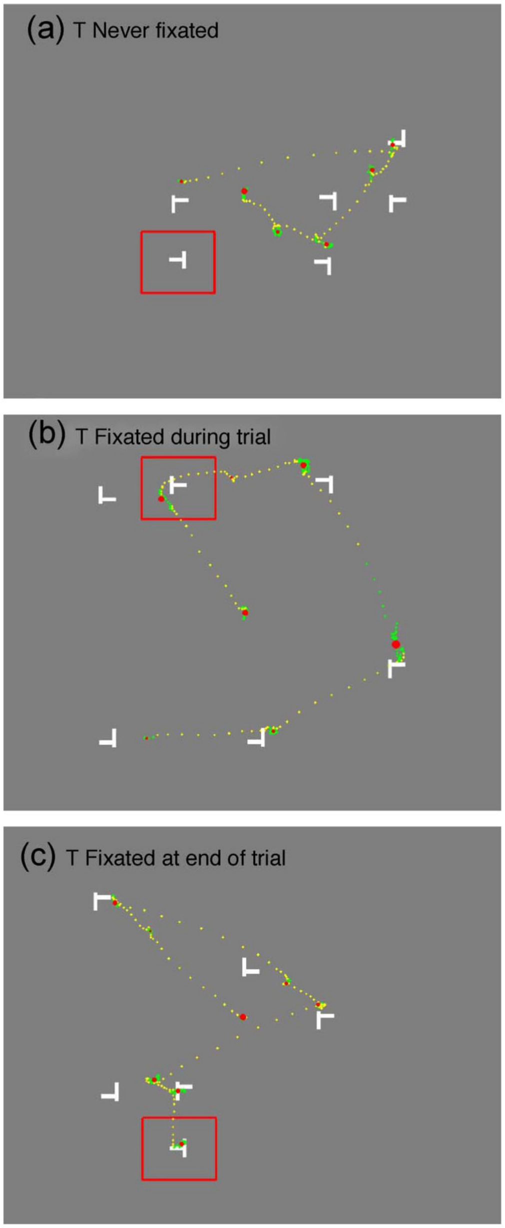

Figure 4.

Example scan paths from low prevalence miss trials classified into trials where (a) the target was never fixated, (b) the target was fixated during the trial, and (c) the target was fixated at the end of the trial. Red dots indicate fixations, green dots identify data points that were indicative of the fixation and yellow dots trace the saccade. The red box denotes the region defined as ‘landing on the target’.

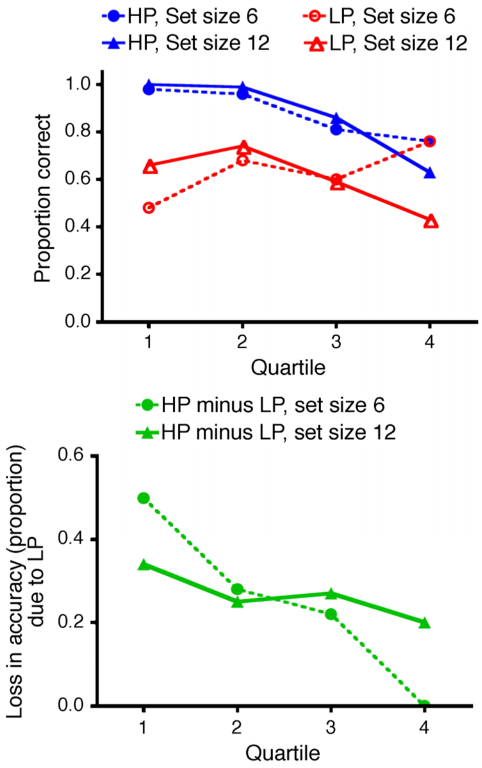

Figure 5.

(a) Proportion of correct target-present trials as a function of RT quartile. (b) Difference in proportion correct target-present trials between high and low prevalence. Positive values indicate increased miss error rate at low prevalence. HP = high (50%) prevalence, LP = low (2%) prevalence.

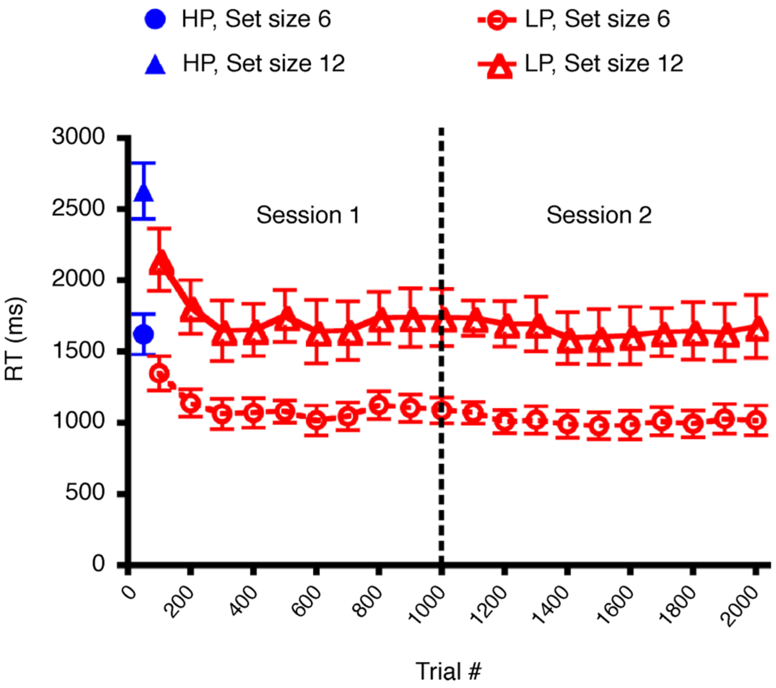

Figure 6.

Mean RT for correct target-absent trials separated by set size in 100 trial bins. Error bars represent one standard error of the mean. Data from the high prevalence condition are shown in blue for comparison. HP = high (50%) prevalence, LP = low (2%) prevalence.

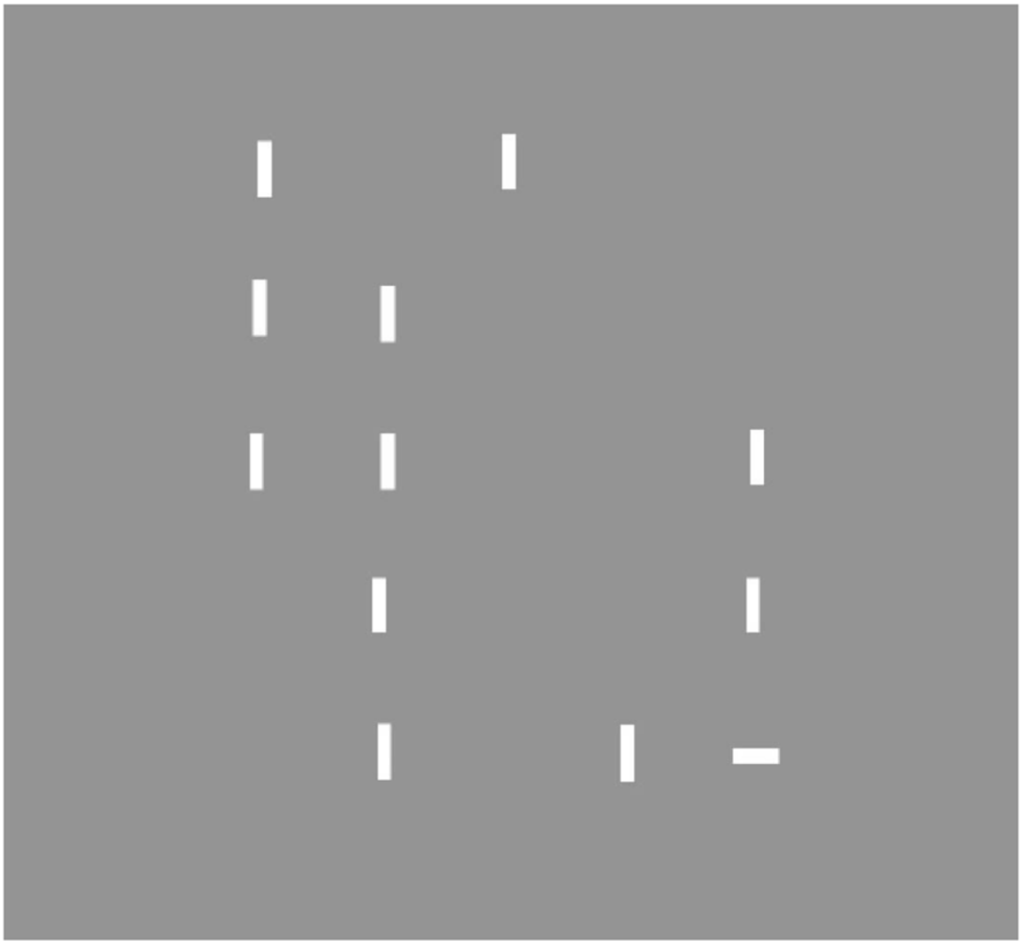

Figure 7.

Sample target-present display for Experiment 2. Participants searched for a horizontal line among vertical lines forming an easy feature search.

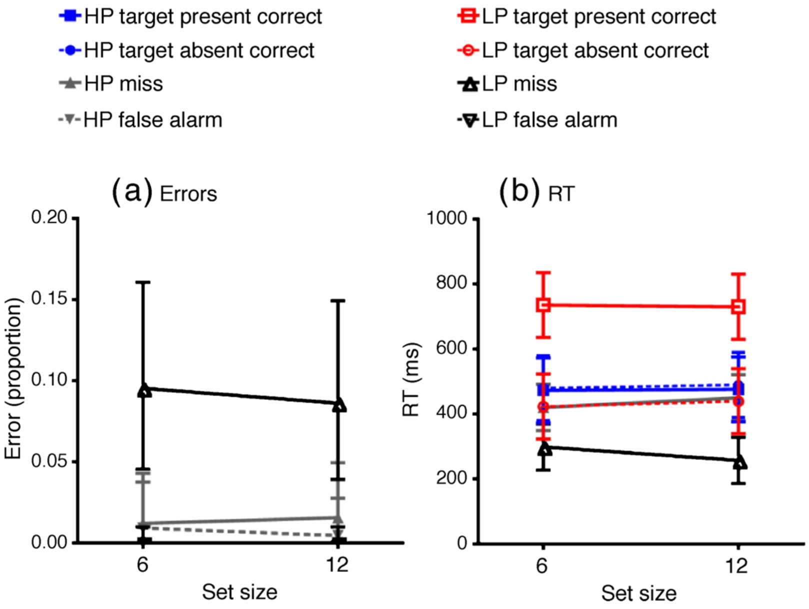

Figure 8.

Results from Experiment 2. Participants searched for a single horizontal line among vertical distractors. (a) Mean error rate (proportion) and (b) mean RT (ms), for correct and incorrect responses for each of the prevalence and target presence conditions, plotted by set size. Error bars represent within-subjects 95% confidence intervals. HP = high (50%) prevalence, LP = low (2%) prevalence.

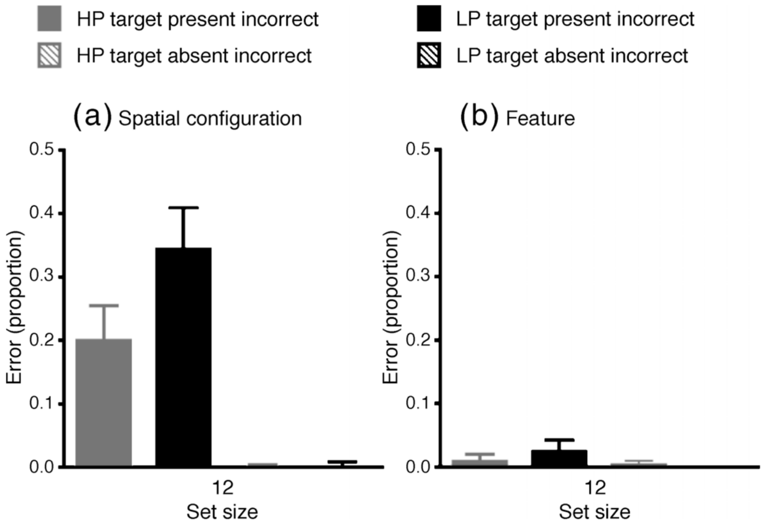

Figure 9.

Mean error rates (proportion) from Experiment 3 where participants were forced to wait a set duration before responding for (a) the spatial configuration search; and (b) the feature search. Error bars represent within-subjects 95% confidence intervals. HP = high (50%) prevalence, LP = low (2%) prevalence.

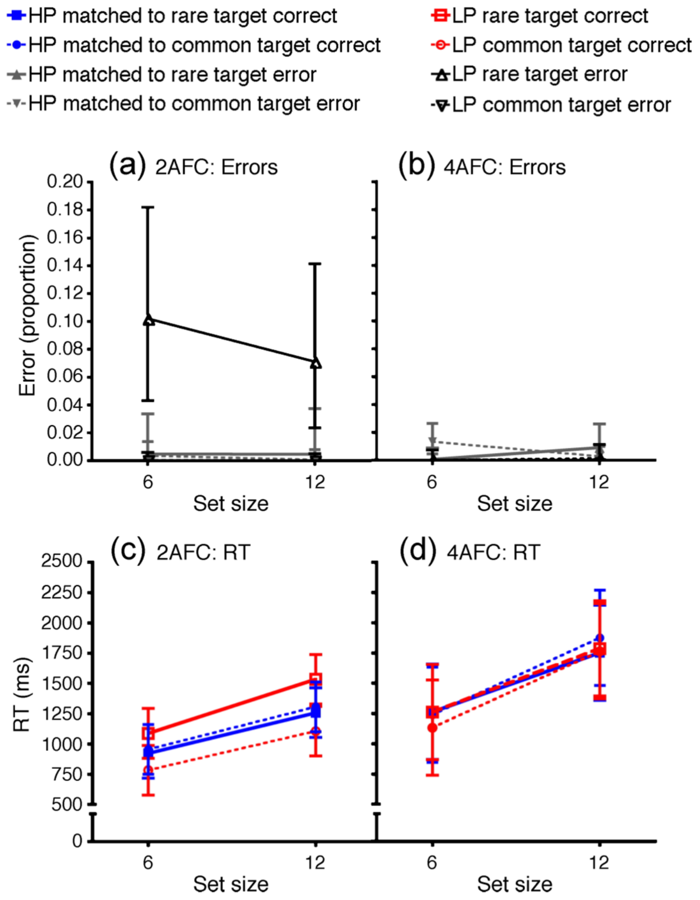

Figure 10.

Results from Experiment 4 where participants identified the orientation of the target T. (a) 2AFC mean error rate (proportion), (b) 4AFC mean error rate (proportion), (c) 2AFC mean RT (ms), and (d) 4AFC mean RT (ms), for correct and incorrect responses for each of the prevalence and target frequency conditions, plotted by set size. Error bars represent within-subjects 95% confidence intervals. AFC = alternative forced-choice. HP = high prevalence, LP = low prevalence.

Table 1

Miss errors classified by fixations on target, summed across participants, collapsed across set-size.

	Proportion of total target-present trials (N misses)		
Fixations on target	**High prevalence**	**Low prevalence**	**Additional errors due to low prevalence (low–high; proportion)**
None	0.06 (39)	0.26 (82)	0.20
During trial	0.04 (27)	0.05 (16)	0.01
End of trial	0.02 (13)	0.08 (25)	0.06