Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-025-03084-9

Reducing the low prevalence effect: Does similarity search translate to binary decisions?

Greer Gillies[1] · **Anna Kosovicheva**[1,2]

Accepted: 11 April 2025

Abstract

In visual search, observers often miss rare targets. This low prevalence effect (LPE) is resistant to many cognitive interventions. However, a recent study showed that having participants identify the item that was most similar to the target (similarity search) eliminated the LPE. As real-world searches often require binary decisions (e.g., is there a threat in this bag?) we tested whether the benefits of similarity search generalize to binary decisions and to more naturalistic stimuli. Participants searched for *T* shapes amongst near-*T* distractors and the prevalence of true *T*s was manipulated. In the similarity-search-only condition, participants clicked on the "most *T*-like object." In the similarity search & binary decision condition, participants additionally reported whether the chosen item was a true *T* (yes/no). We found that in some circumstances, similarity search can be used to attenuate the LPE. However, there was an LPE for the binary decision task. Participants were less likely to classify the target as a true *T* during low prevalence compared with high. We replicated this result in an additional experiment using more naturalistic stimuli. Participants watched clips of road videos and clicked on the "most hazardous location" in the video, followed by a binary decision ("would you need to respond to that hazard? yes/no"). Though participants located the hazards regardless of prevalence, there was an LPE for the binary decision task. Together, these results indicate potential limitations in applying similarity search outside the laboratory; the LPE is still seen in these searches if a binary decision is involved.

Keywords Visual search · Attention

Many real-world visual searches rely on our ability to detect rare but important items (e.g., road hazards, threats in luggage). Despite the importance of the targets, observers often miss them when their prevalence is low. For example, newly trained TSA (Transportation Security Administration) officers are worse at finding target items (e.g., weapons) in simulated X-ray screenings of luggage when those targets are rare (Wolfe et al., 2013). Importantly, this phenomenon—the low prevalence effect (LPE)—has been demonstrated in many naturalistic visual search contexts, including radiological image search (Evans et al., 2013), simulated baggage screening (Wolfe et al., 2013), and the detection of deepfake videos on the internet (Josephs et al., 2024). The LPE can also potentially impact drivers' ability to find certain types of vehicles (Beanland et al., 2014) and ability to detect and respond to road hazards (Kosovicheva et al., 2023). Similar results have been shown in the classic vigilance literature, in which observers respond to infrequent targets that appear over an extended period of time (Mackworth, 1948). Performance decreases as the prevalence of the targets decreases (Baddeley & Colquhoun, 1969; Broadbent & Gregory, 1965; Colquhoun, 1961).

From a signal detection framework, the LPE reflects a shift towards a more conservative response criterion (bias away from making a "target-present" response). Target prevalence rarely impacts sensitivity (d'), suggesting that observers do not just become completely careless during periods of low prevalence (Thomson et al., 2016; Wolfe et al., 2007). Prevalence also impacts the amount of time participants spend searching an array. Participants terminate search earlier during periods of low prevalence compared with high (Wolfe & Van Wert, 2010).

✉ Greer Gillies
greer.gillies@mail.utoronto.ca

1 Department of Psychology, University of Toronto, Toronto, Canada

2 University of Toronto, Mississauga, Canada

Published online: 13 May 2025

Because of the LPE's impact on real-world searches, previous studies have made many attempts to reduce or eliminate the LPE, with mixed success. Some interventions have relied on multiple observers in an effort to catch infrequent targets. For example, Wolfe et al. (2007) tested several interventions using a simulated baggage screening task with either high- or low-prevalence targets (weapons amongst other objects). One variant of the task tested whether having multiple participants view the same images would reduce the LPE. Despite having two sets of eyes on the suitcases, there was still an LPE, though performance was slightly better relative to individual performance. However, more recent work measuring the impact of multiple searchers on the LPE has shown that searching in pairs can substantially reduce miss rates in a radiological image search task, particularly when motor errors are accounted for (Kunar et al., 2021).

Other interventions have aimed to reduce the LPE by modifying individual participants' decision boundaries by changing either the expectation of target prevalence, or the payoffs associated with each response. For example, in the final experiment of Wolfe et al. (2007), participants completed blocks of low-prevalence searches with no feedback, interspersed with high-prevalence "bursts" (i.e., a short block of trials with high target prevalence) with feedback throughout the experiment. This eliminated the LPE, and participants maintained a more liberal criterion during low prevalence. This suggests that observers use feedback to estimate target prevalence.

In line with this finding, several other studies have shown that feedback manipulations can reduce, eliminate, or even reverse the LPE in some contexts. For example, removing response feedback has been shown to reduce the low-prevalence effect in simulated baggage screening tasks (Van Wert et al., 2009) as well as a computer-based road-hazard detection task (Kosovicheva et al., 2023). Removing feedback can also shift participants' responses to become overly liberal when making perceptual decisions about ambiguous stimuli (Lyu et al., 2021). A key factor may be participants' expectations of target prevalence (Cox et al., 2021). For example, another study by Schwark et al. (2012) found that giving incorrect feedback to increase participants' expectations of target prevalence could alter the LPE. If observers attempt to equate the number of misses and false alarms during search, then informing participants they are making more misses than they actually are would lead to participants adopting at more liberal criterion. Indeed, false feedback about miss rate improved target detection during low prevalence.

Despite the effectiveness of removing feedback on reducing or eliminating the LPE, removing or modifying feedback in real life is often not possible as feedback can be self-generated. Consider that when driving, each moment you do not have to brake or swerve to avoid a road hazard is "feedback" that a hazard was not present. Most real-world errors on the road will be false alarms (e.g., braking too early) rather than misses (e.g., getting in a fender bender). Self-generated feedback (e.g., "Oh, I didn't need to brake that early") about those errors may make drivers less likely to respond to actual hazards (which are rare). Therefore, removing feedback is not always a viable option to eliminate the LPE in real-world scenarios. An alternative to modifying feedback may be to modify the payoffs and penalties associated with each type of response (Hadjipanayi et al., 2023). However, similarly, this is not always possible in real-life situations, and the LPE has been shown to still persist in some high-stakes real-world scenarios (Evans et al., 2013).

Finally, in addition to the interventions described above, previous work has examined whether changing some element of the task or the response required of participants would reduce the LPE. For example, one intervention reduced the LPE by eliminating sources of motor error; participants showed a reduced miss rate when they were given an opportunity to correct their initial "target-present" or "target-absent" response in a low-prevalence search task (Fleck & Mitroff, 2007). However, further work with this approach has suggested that this may be less effective for complex searches and may depend on the nature of the feedback provided (Van Wert et al., 2009). Moreover, several studies have shown a persistent LPE with correctable searches (Kunar et al., 2021; Rich et al., 2008).

In another task-based intervention, a recent study showed that the LPE could be eliminated by using a simple cognitive strategy (Taylor et al., 2022). Rather than have observers search for a specific target, participants were instructed to engage in "similarity search"—finding the most target-like object on every trial, regardless of whether or not a target was actually present. Taylor et al. (2022) had participants search for *T* shapes (in which the vertical line segment perfectly bisected the horizontal one) among *T*-like distractors (which varied in the size of the offset between the two line segments), and the prevalence of the true *T*s was manipulated (high = 50%, low = 10%). In one condition, participants reported whether or not a *T* was present ("present/absent search"). In the other condition, participants were asked to locate the most *T*-like object in the array on every trial ("similarity search"). Participants were given feedback on each trial. In the present/absent search condition, there was an LPE for target detection. However, in the similarity search condition, participants were just as good at locating the true *T*s during periods of high prevalence as low, effectively eliminating the LPE. The authors speculate that because similarity search requires a target-present response on each trial, participants maintained their expectations for a target on each trial.

In this study, we expand upon the similarity search method to test its feasibility as an intervention in more naturalistic domains. Though similarity search is easy to carry out, it is not wholly practical. Imagine standing in the TSA line where the officer has to find the most weapon-like item in every bag. Instead of pulling out hairbrushes and curling irons from every piece of luggage, the officer would instead need to ask themselves "is this item I selected from the X-ray an actual weapon?" Following similarity search, observers would ultimately need to make a binary decision about the selected target (e.g., is this item in the luggage actually a weapon? Do I need to hit the brakes?).

In three experiments, we tested whether similarity search extends to these sorts of binary decisions. In other words, can participants correctly report whether or not they selected the "true" target? One possibility is that similarity search shifts observers' response criterion in a way that generalizes across multiple decision types; in other words, they would find the item that most closely resembles a target *and* correctly classify the true targets when they are present. Alternatively, observers may set a separate criterion for these binary decisions and fail to correctly classify the target, even if they find it in the similarity search task. In the first two experiments, we tested this by having participants report whether a target was present after each similarity search trial. In a third experiment, we test whether similarity search is effective with more naturalistic stimuli by having participants identify the most hazardous object in videos of real road scenes.

Experiment 1

In Experiment 1, participants performed a search task in which they looked for *T*s amongst *T*-like distractors. We compared observers' error rates in two different conditions: a "similarity-search-only" condition and a "similarity search & binary decision" condition. For the "similarity-search-only" condition, participants searched for most *T*-like object out of a display of shapes varying from perfect *T*s to *L*s. The prevalence of true *T*s varied between two prevalence conditions (low, 10% target prevalence; high, 50% target prevalence). Participants made their decisions by clicking on the shape they thought was the target. After each click, participants were provided feedback by showing where the correct target was for that trial. For the "similarity search & binary decision" condition, participants searched for the most *T*-like object, but after clicking it, they then reported whether or not the selected object was a true *T*. Participants were given feedback based on the accuracy of their binary decision.

If the benefits of similarity search transfer to binary decisions, we would expect to see that participants would successfully classify their selected targets, regardless of the prevalence of the true *T*s. Conversely, if similarity search does not transfer to binary decisions, there may be a secondary LPE for the binary decision task (i.e., participants will be worse at classifying the true *T*s as true *T*s during periods of low prevalence).

Methods

Participants

Participants were recruited using Prolific (https://www.prolific.com). Prolific is an on-demand self-service data-collection platform. Each participant provided electronic consent to the protocol approved by the Research Ethics Board of the University of Toronto prior to participation and received monetary compensation for their participation (5 GBP for Session 1 and 5 GBP for Session 2 plus an 8 GPB "completion bonus"). All participants self-reported were fluent in English, had normal or corrected-to-normal vision, and were from the USA, Canada, or the United Kingdom.

A total of 34 participants were recruited, and none were excluded (see "Participant Exclusion Criteria" below for details). Power calculations, based on an effect size from a pilot study with 20 participants (Cohen's $d = 0.5$), indicated that a minimum of 34 participants were necessary to detect a significant difference in miss rate during low prevalence based on task type (similarity search & binary decision, similarity search only) at 80% power. Power calculations were conducted using G*Power 3 (Faul et al., 2007).

The mean age of the sample was 29.4 years (range: 22–36) with 19 men and 15 women.

Participant exclusion criteria The preregistered exclusion criteria were based on accuracy during the high-prevalence condition for the similarity-search-only task. The preregistration can be found at (https://osf.io/rszkg). If a participant had an accuracy of less than 30% in the high-prevalence condition for the similarity-search-only task, their data were to be excluded. No one was excluded based on these criteria.

Apparatus

All data were collected online. Participants were directed from Prolific to Qualtrics (https://www.qualtrics.com), where they read and digitally signed a consent form. After signing, they were redirected to Pavlovia (Peirce et al., 2019). The experiment was coded using Psychopy3 (Peirce et al., 2019). Participants were only permitted to do the experiment on a desktop or laptop computer.

Stimuli

The stimuli included a set of "true-*T*" and "near-*T*" shapes (see Fig. 1). A true *T* is a shape where the stem perfectly bisects the branch of the *T* at the midpoint. For near *T*s, the

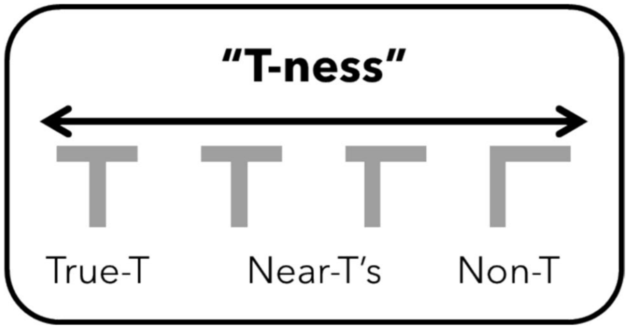

Fig. 1 Sample stimuli for Experiments 1 and 2. The targets were either true *T*s or near *T*s. For true *T*s, the stem bisected the branch at the midpoint. The near *T*s varied in how much the stem was offset from the branch

stem is offset from the center of the branch. Each shape was generated from horizontal and vertical line segments that filled a square region that was resized for each participant to fill 5% of the screen's height. Shapes had a stroke width-to-shape-height ratio of 1:5.5. The smallest offset from the middle of the branch for the "near *T*s" was 4.5% of the shape height, and the largest offset from the middle of the branch was 41% of the shape height (a perfect *L* shape). Shape image files were pregenerated in MATLAB according to two parameters: offset (0% for true *T*s, 4.5% through 41% of shape height [in increments of 2.27%] for near *T*s), and shade of grey (RGB value 100 through 160). Each shape was produced in each shade of grey (i.e., 61 grey levels per shape).

The 16 *T* shapes (one target, 15 distractors) were presented in random locations in an 8×8 square grid (the grid was 70% × 70% of the screen's height, each position was 10% of the screen's height apart) on a dark-grey background (RGB value of 40), with a horizontal and vertical positional jitter (maximum of 2.5% of the screen's height) and a random rotation angle (chosen from 0, 90, 180, or 270 degrees) applied to each shape.

Procedure

This similarity search task was adapted from Experiment 3 in Taylor et al. (2022). On each trial, participants were asked to search for the most *T*-like shape amongst 15 *T*-like distractors (see Fig. 2). Participants indicated their response by clicking on the shape. Participants were given a maximum of one minute to search the array (there were no participants who reached this time limit). The prevalence of the true *T*s was manipulated. During high prevalence, a true *T* was present 50% of the trials. During low prevalence, it was 10%. On target-present trials, the display consisted of one true-*T* target and 15 near-*T* distractors. On trials where a true *T* was not present, the display consisted of entirely near-*T* shapes, and the "target" for that trial was the shape that was closest to a true *T* (i.e., had the smallest offset among the shapes on the screen). Displays were generated by first drawing fifteen near-*T* distractors at random. Then, a range of possible offset values for the target near *T* was calculated (e.g., if the smallest offset in the distractor list was 18%, the near-*T* target for that trial could have an offset between 4.5% and 13.5%). The near-*T* target offset was then randomly selected from this range of possible offsets, resulting in exactly one unique target per trial.

There were 50 trials in the high-prevalence condition, and 250 trials in the low-prevalence condition. Order of prevalence was counterbalanced between participants. Prior to the start of each trial block, participants performed 20 practice trials (40 practice trials in total), and participants were informed about the prevalence of the true *T*s. Participants were given an optional 1-min break every 50 trials.

Similarity search only In the similarity-search-only task, after a participant selected a shape, they were given feedback based on the accuracy of their target selection. If they were incorrect, a red box was shown around the correct shape. If they were correct, a green box was shown around the correctly identified target shape. The feedback was on-screen for 2 s. Participants were only provided with feedback about the accuracy of their choice, but they were not told whether-or-not the selected target was a true-T.

Similarity search and binary decision In the similarity search & binary decision task, after a participant selected a shape, they were then asked, "Was that image a true *T*? Yes or no." Participants indicated their responses with their keyboards ("up" for yes and "down" for no). They were then given feedback (correct/incorrect) based on the accuracy of their classification. The feedback was on screen for two seconds. It is important to note that participants were given feedback based on the binary decision, not the search task. For example, if a participant selected one of the distractors on a true-*T*-present trial, but they said their selected target was a non-*T*, they would be told they were "correct." Therefore, they were not given information about the accuracy of the initial target selection during the search task.

The two tasks (similarity search only and similarity search & binary decision) were completed across two sessions on different days, with task order counterbalanced across participants.

Analysis

Similarity search We compared accuracy on the similarity search task (correctly identifying the location of the true *T* when it was present) across the two prevalence conditions

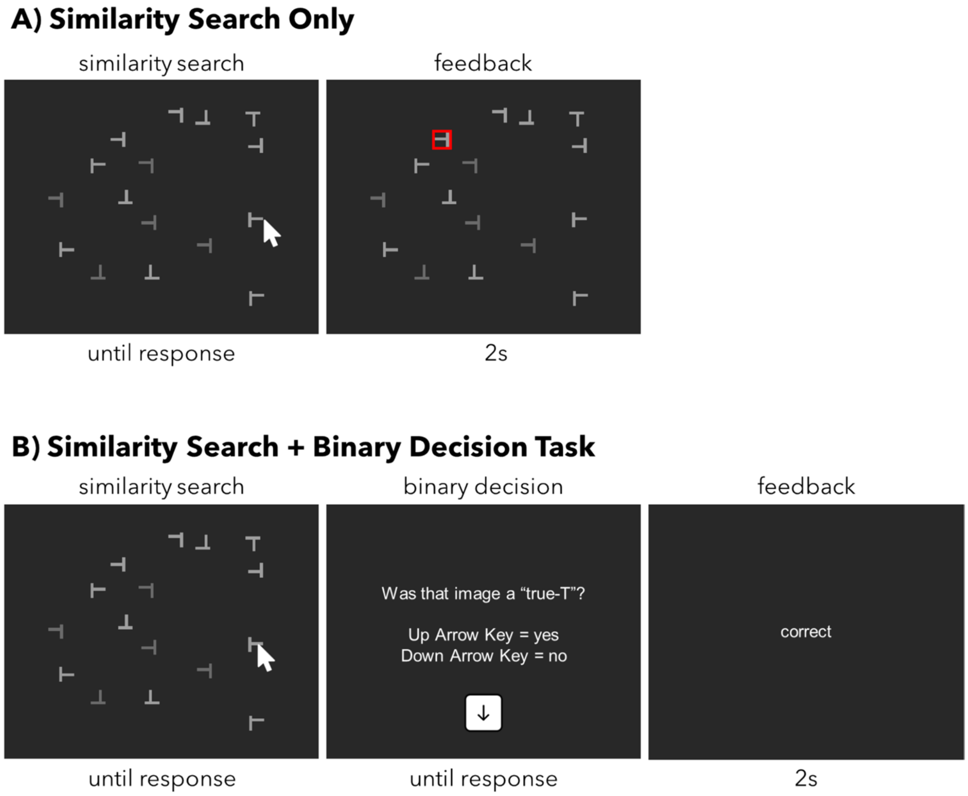

Fig. 2 Trial Sequence for Experiments 1 and 2. **A)** The similarity-search-only task. Participants were told to click on the most *T*-like shape and were then given feedback based on their accuracy. **B)** The similarity search & binary decision task. Participants were told to click on the most *T*-like shape and were then asked whether or not it was a true *T*. Participants were then given feedback on the accuracy of their classification. In Experiment 1, the prevalence for the true *T*s was 10% (low) and 50% (high). In Experiment 2 it was 4% (low) and 50% (high)

(low, high) and task (similarity search only and similarity search & binary decision) using a 2 × 2 repeated-measures analysis of variance (ANOVA). Critically, we examined the interaction between task and prevalence. We also performed two paired-sample *t* tests to examine the difference in accuracy between high and low prevalence for both task types.

We also examined the accuracy of the true-*T*-absent trials to examine whether participants were successfully locating the near-*T* targets using a 2 (prevalence: low, high) × 2 (task: similarity search only and similarity search & binary decision) repeated-measures ANOVA.

We also examined average response times across the high- and low-prevalence conditions on the true-*T*-present trials (correct responses only) for both tasks using two paired-sample *t* tests.

Binary decision For each participant and prevalence condition, we calculated the miss rates out of the set of trials in which the participant correctly clicked on the true *T* on the search task. The miss rate was the proportion of true-*T*-present trials in which the participant successfully located the true *T* on the search task and then did not classify it as a true *T*. (In this case, a "hit" is classified as a trial in which the participant successfully located the true *T* on the search task and correctly classified it as a true *T*, so hit rate is equal to 1 minus the miss rate as in signal detection theory.)

We also calculated the false-alarm rate out of the set of trials in which the participant clicked on a non-*T* shape in the search task. The false-alarm rate was the proportion of these trials in which the participant classified the non-*T* shape they had clicked as a true *T*. As these responses are coded based on the item the participant clicked rather than the full set of items in shown on a given trial, a false alarm could therefore occur on a true-*T*-present *or* a true-*T*-absent trial.

We then examined differences in misses and false alarms across the prevalence conditions using two paired-sample *t*-tests.

In addition, we performed a signal detection analysis to examine changes in sensitivity (d') and criterion (measured as bias; b) on the binary decision task using a two paired-sample *t* tests. Bias and d' were both calculated using the Psycho library in R (Makowski, 2018).

Results

Similarity search

Figure 3 shows mean accuracy (percentage of correct clicks on true-*T*-present trials) for each condition in the similarity search task. The 2 (task: similarity search only, similarity search & binary decision) × 2 (prevalence: low, high) repeated-measures ANOVA revealed no main effect of task on accuracy, $F(1, 33) = 2.67$, $p = .11$, $\eta_p^2 = .08$. There was a significant main effect of prevalence $F(1, 33) = 30.34$, $p < .001$, $\eta_p^2 = .48$. Specifically, participants were significantly less accurate at locating the true *T*s during periods of low prevalence than high. There was no interaction between task and prevalence, $F(1, 33) = 1.21$, $p = .28$, $\eta_p^2 = .04$.

We also confirmed that the LPE was present in each of the two task types using two paired-sample *t* tests (corrected for multiple comparisons, $\alpha = .025$). Participants were significantly worse at finding the true *T*s in the similarity-search-only task, $t(33) = 4.74$, $p < .001$, Cohen's $d = .81$, and in the similarity search & binary decision task, $t(33) = 4.05$, $p < .001$, Cohen's $d = .70$.

For the true-*T*-absent trials, there was a main effect of task, $F(1, 33) = 11.61$, $p = .002$, $\eta_p^2 = .26$. Specifically, participants were significantly better at finding the near-*T* targets on the similarity-search-only task compared with the similarity search & binary decision task. There was also a main effect of prevalence (prevalence relating to the prevalence of the true *T*s), $F(1, 33) = 12.67$, $p = .001$, $\eta_p^2 = .28$. Participants were significantly better at finding the near-*T* targets during high true-*T* prevalence compared with low. The interaction was not significant, $F(1, 33) = 1.91$, $p = .18$, $\eta_p^2 = .06$.

Reaction time For the similarity-search-only task, there was no significant difference in average response times between high and low prevalence (4.12s vs. 3.90s), $t(33) = .99$, $p = .33$, Cohen's $d = .17$.

For the similarity search & binary decision task, there was no significant difference in response times between high and low prevalence (4.49 s vs. 4.33 s), $t(33) = .69$, $p = .49$, Cohen's $d = .12$.

Binary decision Figure 4 shows the miss rate, false-alarm rate, *d′*, and criterion for the binary decision task. Miss rate was significantly greater in the low-prevalence condition compared with high, $t(33) = 3.22$, $p = .003$, Cohen's $d = .55$. There were also significantly fewer false alarms in the low-prevalence condition compared with high, $t(33) = 6.74$, $p < .001$, Cohen's $d = 1.16$.

There was no significant difference in *d′* between the low- and high-prevalence conditions, $t(33) = .97$, $p = .34$, Cohen's $d = .17$. However, consistent with the pattern of miss rates and false alarms, participants were significantly more conservative (had a more strict criterion) in their willingness to say a selected target was a *T* in low prevalence than in high, $t(33) = 6.28$, $p < .001$, Cohen's $d = .21$.

Discussion

There appears to be a double low-prevalence effect. During periods of low prevalence, participants were less likely to locate the true *T* on the search task compared with high prevalence.

In addition, even when they *did* locate the true *T* on the search task during low prevalence, they then failed to

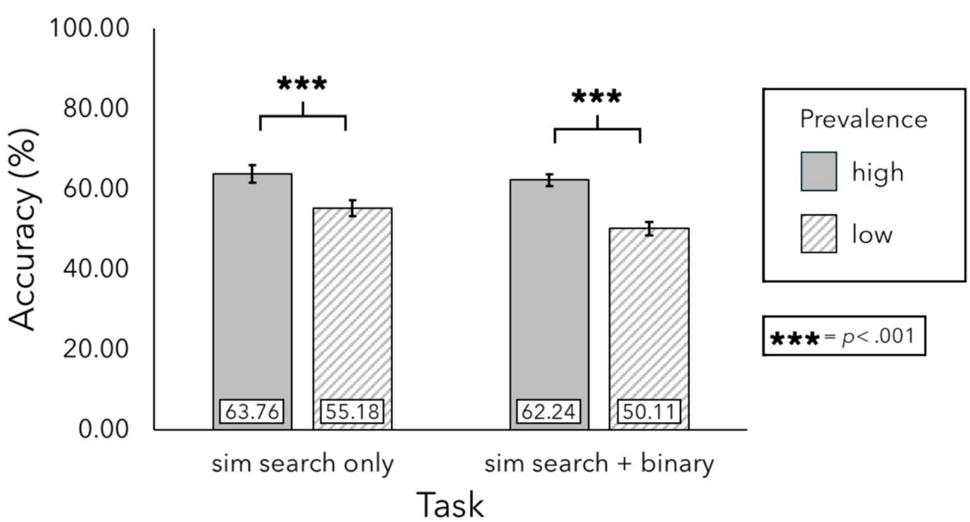

Fig. 3 Similarity search accuracy for Experiment 1. Participants were significantly worse in locating the true *T*s during low prevalence compared with high for both task types. Error bars are Morey's standard error of the mean (*SEM*; Morey, 2008)

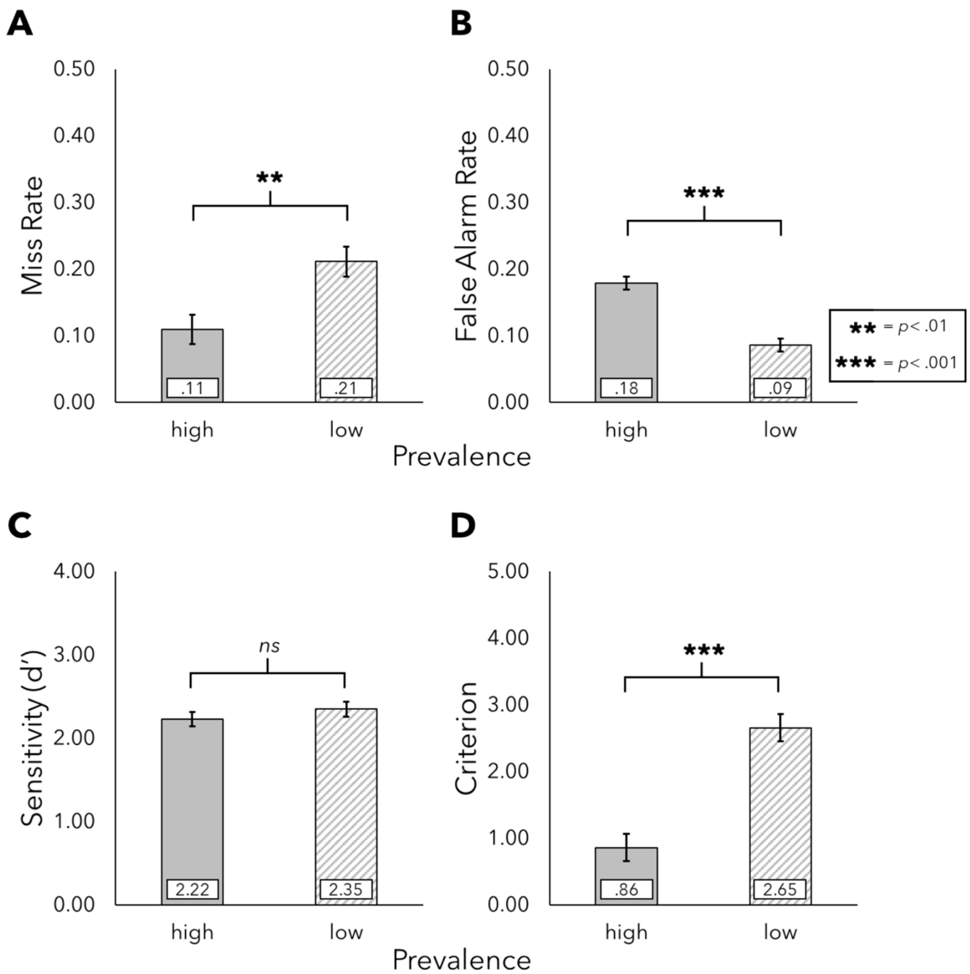

Fig. 4 LPE for the binary decision task in Experiment 1. **A** Misses were significantly higher in the low prevalence condition compared with high. **B** False alarms were significantly lower in the low prevalence condition compared to high. **C** Sensitivity was not impacted by prevalence. **D** Criterion was significantly higher (more conservative) in the low-prevalence condition compared with high. Error bars are Morey's *SEM* (Morey, 2008)

classify it as a true *T* on the binary decision task, with nearly twice the proportion of misses in the low-prevalence condition compared with the high-prevalence condition (21% vs. 11%). In sum, similarity search did not provide any benefit on the binary classification task.

Experiment 2

In Experiment 1, we observed an LPE in the similarity search task in both conditions (similarity search only and in the similarity search & binary decision conditions), with no interaction between prevalence and condition. Furthermore, we observed an LPE for binary decisions. Although we observed a significant LPE when observers engaged in a similarity search task, which was a manipulation intended to eliminate the LPE, we note that the LPE was not very large in this case (similarity search only: 63.75% target detection for high, 55.18% for low; similarity search & binary decision: 62.24% for high and 50.12% for low). These values are also comparable with the similarity search performance reported by Taylor et al. (2022).

One possible explanation for this result is that the benefits of similarity search may be more easily observed at lower prevalence rates. Therefore, we carried out an additional experiment and exaggerated the difference between high and low prevalence from 50% and 10% to 50% and 4%.

Methods

Participants

Participants were recruited and screened in the same manner as in Experiment 1.

A total of 43 participants were recruited, and three were excluded (see "Participant Exclusion Criteria"). Power calculations, based on an effect size from a pilot study with 20 participants (Cohen's $d = 0.5$), indicated that a minimum of 34 participants are necessary to detect a significant difference in miss rate during low prevalence based on task type (similarity search & binary decision, similarity search only) at 80% power. Power calculations were conducted using G*Power 3 (Faul et al., 2007).

The mean age of the final sample was 31.32 years (range: 22–36), with 33 men and seven women.

Participant exclusion criteria We used the same exclusion criteria as in Experiment 1. Three participants were excluded based on these criteria. See (https://osf.io/kqbrm) for the preregistration.

Stimuli and procedure

The apparatus, procedure, stimuli, and analyses were identical to those used in Experiment 1, except for the following: In the low-prevalence condition, the prevalence of the true *T*s was reduced to 4%. There were 300 trials in the low-prevalence condition.

Results

Similarity search

The 2 (task: similarity search only, similarity search & binary decision) × 2 (prevalence: low, high) repeated-measures ANOVA revealed no main effect of task on accuracy on true-*T*-present trials, $F(1, 39) = .03$, $p = .86$, $\eta_p^2 = 8.03\times10^{-4}$ (see Fig. 5). There was a significant main effect of prevalence $F(1, 39) = 17.99$, $p < .001$, $\eta_p^2 = .32$. Specifically, participants were significantly less accurate at locating the true *T*s during periods of low prevalence than high. There was no interaction between task and prevalence, $F(1, 39) = 3.76$, $p = .06$, $\eta_p^2 = .09$.

We also examined whether the LPE was separately present in each of the two task types using two paired-sample *t* tests (corrected for multiple comparisons, $\alpha = .025$). There was no significant difference in locating the true *T*s during low prevalence compared with high for the similarity-search-only task, $t(39) = 1.82$, $p = .077$, Cohen's $d = .29$. Conversely, participants were significantly worse at locating the true Ts during periods of low prevalence compared high for the similarity search & binary decision task, $t(39) = 4.71$, $p < .001$, Cohen's $d = .74$. This comparison suggests that similarity search may serve to reduce or eliminate the LPE in certain contexts.

For the true-*T*-absent trials, there was a main effect of task, $F(1, 39) = 4.94$, $p = .03$, $\eta_p^2 = .11$. Specifically, participants were significantly better at finding the near-*T*-targets on the similarity-search-only task compared with the similarity search & binary decision task. There was also a main effect of prevalence (prevalence relating to the prevalence of the true *T*s), $F(1, 39) = 9.64$, $p = .004$, $\eta_p^2 = .20$. Participants were significantly better at finding the near-*T* targets during high true-*T* prevalence compared with low. The interaction was not significant, $F(1, 39) = .004$, $p = .95$, $\eta_p^2 = 1.04\times10^{-4}$.

Reaction time For the similarity-search-only task, there was no significant difference in response times between high and

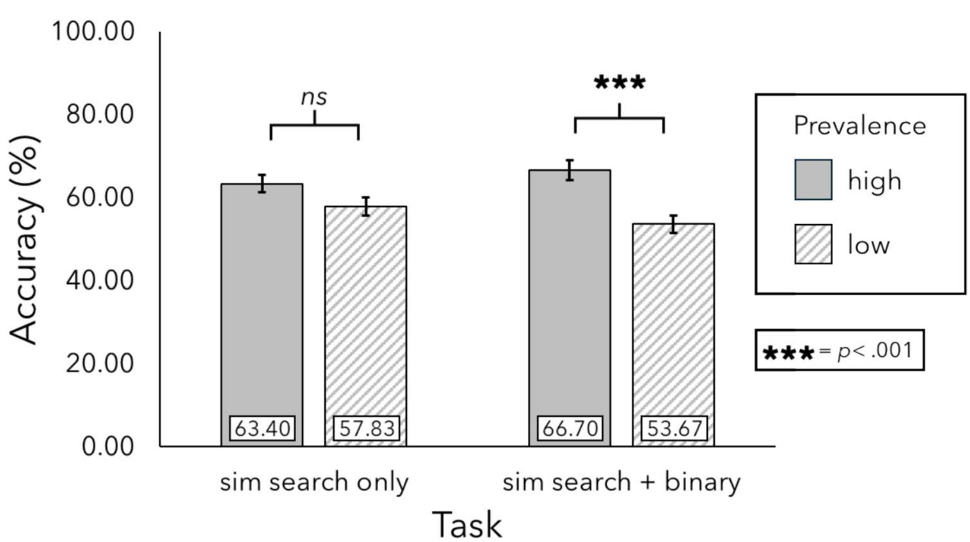

Fig. 5 Similarity search accuracy for Experiment 2. In the similarity-search-only task, participants were equally good at locating the true *T*s during high and low prevalence. In the similarity search & binary decision task, participants were significantly worse at locating the true *T*s during low prevalence compared with high. Error bars are Morey's *SEM* (Morey, 2008)

low prevalence (4.50s vs. 4.63 s), $t(39) = .42$, $p = .678$, Cohen's $d = .18$.

For the similarity search & binary decision task, there was no significant difference in response times between high and low prevalence (4.79s vs. 4.93 s), $t(39) = .31$, $p = .76$, Cohen's $d = .19$.

Binary decision

Figure 6 shows the miss rate, false-alarm rate, d', and criterion for the binary decision task. Miss rate was significantly higher in the low-prevalence condition compared with high, $t(39) = 4.96$, $p < .001$, Cohen's $d = .78$. There were also significantly fewer false alarms in the low-prevalence condition compared with high, $t(39) = 6.83$, $p < .001$, Cohen's $d = 1.08$.

There was no significant difference in d' between the low- and high-prevalence conditions, $t(39) = .11$, $p = .91$, Cohen's $d = .02$. However, participants were significantly more conservative (had a more strict criterion) in their willingness to say a selected target was a *T* in low prevalence than in high, $t(39) = 5.17$, $p < .001$, Cohen's $d = .82$.

Discussion

Overall, we found a similar pattern of results to that of Experiment 1. When similarity search was combined with a binary decision task, we observed a significant low prevalence effect in participants' responses when clicking most *T*-like target. In addition, when asked to classify whether the selected target was a true *T*, participants missed a larger proportion of targets under low prevalence compared with high prevalence (23% vs. 8%). When participants completed

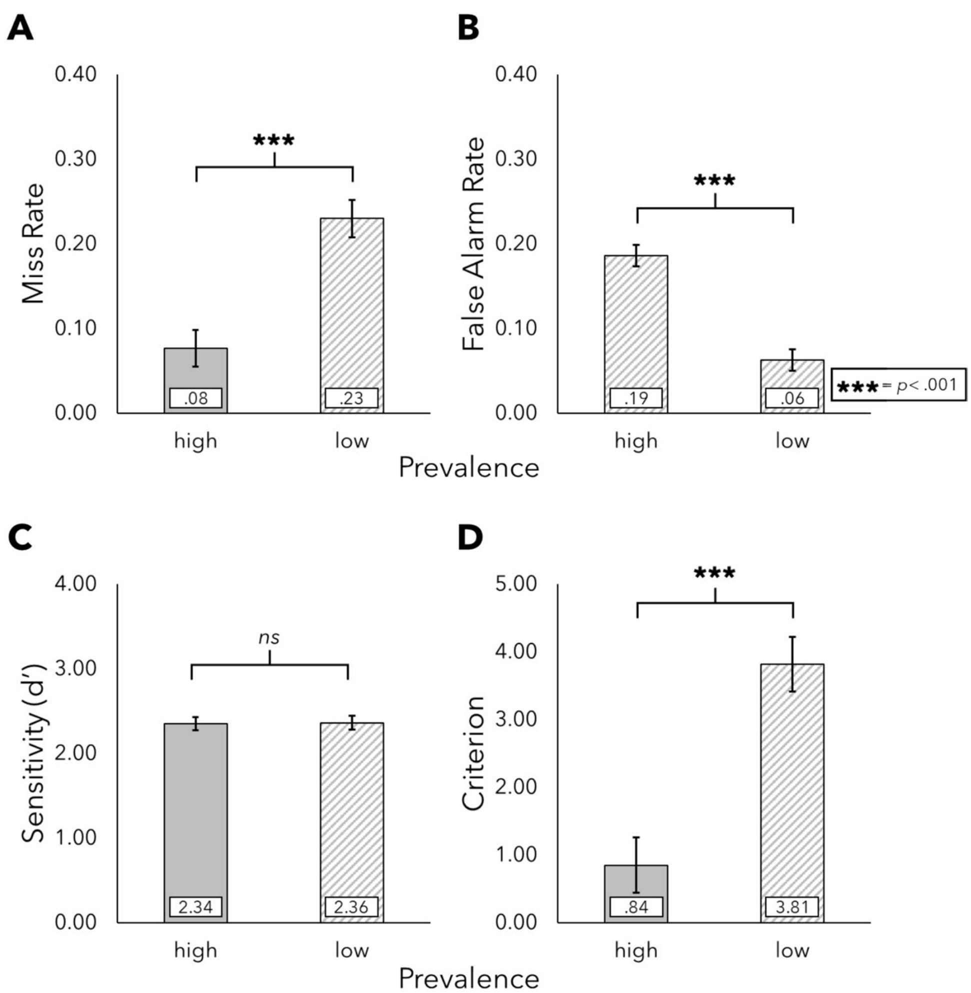

Fig. 6. LPE for the binary decision task in Experiment 2. **A** Misses were significantly higher in the low prevalence condition compared with high. **B** False alarms were significantly lower in the low prevalence condition compared with high. **C** Sensitivity was not impacted by prevalence. **D** Criterion was significantly higher (more conservative) in the low prevalence condition compared with high. Error bars are Morey's *SEM* (Morey, 2008)

the similarity search task on its own, however, performance in identifying the True T was not significantly worse in the low-prevalence condition compared with the high-prevalence condition, which replicates previously reported results (Taylor et al., 2022).

Previous work has suggested that interventions to reduce the LPE may vary, depending on the complexity of the task and the nature of the stimuli (Van Wert et al., 2009). Although we still observed an LPE in both task conditions of Experiment 1, and in the similarity search & binary decision condition of Experiment 2, it is possible that the LPE may be entirely absent when similarity search is used with other types of stimuli. We tested this in Experiment 3 and additionally measured participants' performance in making binary decisions following these similarity judgments.

Experiment 3

In Experiment 3, we examined whether any benefits of similarity search are seen with more naturalistic stimuli. Previous work has shown a low-prevalence effect for the detection of hazardous road events when participants viewed videos recorded from dashboard cameras (Kosovicheva et al., 2023). Specifically, we examined participants' error rates on a road-hazard-detection task using the same video set, where participants were engaged in similarity search. When watching road-hazard videos, participants were asked to click on the most hazardous object in the video, regardless of whether a hazard was present. After clicking on the most hazardous object, participants were then asked if the hazard they selected would require a response on the road (e.g., turning the steering wheel, braking). Participants were given feedback based on the accuracy of their binary decision.

Methods

Participants

Participants were recruited using Prolific (https://www.prolific.com). Each participant provided electronic consent to the protocol approved by the Research Ethics Board of the University of Toronto prior to participation and received monetary compensation for their participation identical to Experiments 1 and 2. All participants self-reported they had driver's licenses, were fluent in English, had normal or corrected-to-normal vision, and were from the USA, Canada, or the United Kingdom.

A total of 16 participants were recruited, and none were excluded (see "Participant Exclusion Criteria" below for details). Power calculations, based on an effect size from a pilot study with 10 participants (Cohen's $d = 0.9$), indicated that a minimum of 10 participants was necessary to detect a significant difference in miss rate based on prevalence (low, high) at 80% power. Power calculations were conducted using G*Power 3 (Faul et al., 2007). This final sample size of 16 is above the target sample but is the same size as previous research (Kosovicheva et al., 2023). The mean age of the sample was 29.94 years (range: 23–36) with 10 men and six women.

Participant exclusion criteria The preregistered exclusion criteria were based on participants' performance on catch trials (see Procedure). Preregistration can be viewed online (https://osf.io/khry6). Participants needed to obtain at least 85% accuracy on the catch trials for inclusion in the final sample. No participants were excluded from this experiment.

Apparatus

Apparatus was the same as in Experiments 1 and 2.

Stimuli

Stimuli were composed of road videos from the Road Hazard Stimulus Set (Wolfe et al., 2020), which has been previously used to demonstrate a low prevalence effect for hazard detection in road video (Kosovicheva et al., 2023; see Fig. 7). The stimulus set contains a wide variety of dashboard-camera videos of real road scenes, sourced from the internet, that contain hazardous events, along with a matched control set of nonhazardous video clips, recorded across a range of different environments (e.g., highway, city streets), weather, and lighting conditions. Here, a hazard is defined as a situation requiring an immediate driver response (e.g., turning the steering wheel or braking) to avoid a collision.

Fig. 7 Sample Stimuli for Experiment 3. **A)** Samples from the hazard present videos, image shows the hazard onset. **B)** Samples from the hazard absent videos

The videos were previously annotated for the time that the hazard onset began (first time point at which the hazard deviated from its normal state) and when the driver made a response (first time point at which the response—braking or swerving—was visible in the video). For the hazard-present videos, the location of the hazard (e.g., a car swerving into a lane) was also annotated (for details, see Song & Wolfe, 2024). The annotated hazard location was defined by a polygon, annotated for each frame between the hazard onset and the time of the driver response. For analyzing click locations in the current experiment, hazard location was determined by the location of the hazard at the time of the driver response, which is the last frame of the video shown to participants.

For the hazard-present clips, the experiment used 201 of the original videos from Wolfe et al. (2020). All of the original videos were trimmed into 333-ms long clips for the purposes of this experiment. For the hazard present videos, the clips were taken from 333 ms immediately preceding the onset of the driver response to the hazard in the video. For the matched set of hazard absent videos, video segments were taken from the hazard-present videos at least 10 s before hazard onset, whenever possible. From these segments, we extracted 333-ms clips every 3 s in the video. With this video set and durations tested, the hazard and nonhazard videos are readily discriminable at 87% accuracy (Kosovicheva et al., 2023).

In the final video set, there were 201 hazard present videos (with annotations for hazard onset and location), and 924 hazard absent videos (no annotations available). Each video had a resolution of 1,280 × 720 pixels and a frame rate of 30 fps.

For the catch trials, we used 1,000 ms video clips of non-driving scenes, sourced from YouTube and public-domain stock footage websites (e.g., Pexels.com). There were 80 total catch-trial videos, composed of footage of everyday settings (e.g., a living room, a beach) and activities (e.g., cooking, hiking). Each video had a resolution of 1,280 × 720 pixels. These trials were primarily used as a data quality measure and to remove possible automated (i.e., nonhuman responses).

Procedure

This procedure was adapted from Kosovicheva et al. (2023). Each trial started with a random noise mask (1,280 × 720 pixels) for 250 ms (see Fig. 8). Following the mask was a 333-ms road video. After the video, another 250-ms noise mask was presented. In the high-prevalence condition, road hazards were present in 50% of the videos. In the low-prevalence condition, road hazards were present on 4% of trials. Videos were presented in random order, without repeating any videos within a participant. Participants were told what the relative prevalence of the hazards would be ("hazards will be relatively rare" or "hazards will appear frequently") prior to the start of the experiment. For the low-prevalence condition, there were 350 trials (14 hazard present trials) and for the high-prevalence condition there were 308 trials (154 hazard present trials). On each trial, participants completed two tasks.

Hazard localization task The first is the "hazard localization" task (analogous to the similarity search task from Taylor et al., 2022). Participants were instructed that they needed to determine the location of the most hazardous object in the video (i.e., what in the video has the greatest potential to be a road hazard), regardless of whether or not a hazard was actually present. After the final mask,

Fig. 8 Trial Sequence for Experiment 3. For the hazard localization task, participants were instructed to click where they thought the most hazardous location in the video was (a 250-ms mask appeared before and after the video). For the binary decision task, participants indicated if they would need to make a response to the hazard they selected in the localization task (brake or turn the steering wheel). Participants were then shown feedback based on the binary decision (regardless of accuracy on the localization task)

participants were presented with a grey rectangle (1,280 × 720 pixels) and were asked to click within the rectangle to indicate where they thought the most hazardous object from the road video was. Responses were not speeded.

Binary decision task The second task was the "binary decision" task, where the participant indicated if they would need to make a response to the hazard they selected in the localization task. More specifically, if the participant was the driver, would that hazard require a response in an on-road situation (e.g., turning the steering wheel or needing to brake) to avoid the hazard? Participants made their decisions using their keyboards (down for "yes, I would need to respond"; up for "no, I would not need to respond"). They were given feedback on every trial as to whether or not their binary decision was correct (regardless of where the participant clicked).

The two conditions (high and low prevalence) were completed across two different sessions on different days, and condition order was counterbalanced across participants. Prior to the start of the main experiment, participants completed 28 practice trials where hazard prevalence was always 50%.

Catch trials Interleaved throughout the experimental trials are catch trials. On a catch trial, a nondriving scene video clip (1,000 ms) was presented, and participants needed to indicate if the scene was indoors (press the "up" key) or outdoors (press the "down" key). There were 28 catch trials during both the high- and low-prevalence conditions, for a total of 56 catch trials across the two sessions.

Analysis

Hazard localization task To examine accuracy on the hazard localization task, we determined if a click was within the hazard location polygon (for hazard-present trials only). Because participants were clicking on a grey rectangle and not the final frame of the video, we expanded the boundary of the hazard polygon by 50 pixels in all directions. This was also done to account for positional error due to representational momentum (Freyd & Finke, 1984) or motor response error. Clicking within the polygon counted as a "correct" response. We then compared accuracy between the high- and low-prevalence conditions using a paired-sample t test.

To show that participants were not just randomly clicking on the screen, we repeated the above analysis but randomly shuffled the polygon coordinates with the videos (so each video now had the wrong hazard location annotation). If participants were just randomly clicking on the screen, shuffling the hazard locations would not change the results.

Binary decision task For each participant and prevalence condition, we calculated the miss rate (proportion of hazard-present trials where the participant successfully located the hazard in the localization task and then said they would not need to respond to that hazard) and false-alarm rate (proportion of hazard absent trials where the participant incorrectly classified the road video as containing a hazard). We then examined differences in hits and false alarms across the prevalence conditions using two paired-sample t tests.

In addition, we performed a signal detection analysis to examine changes in sensitivity (d') and criterion on the binary decision task using a series of paired-sample t tests. Criterion and d' were both calculated using the psycho library in R (Makowski, 2018).

Results

Hazard localization task

There was no difference in accuracy (proportion of correct clicks inside the hazard location polygon) between the high- and low-prevalence conditions, $t(15) = .88$, $p = .40$, Cohen's $d = .22$ (see Fig. 9). Therefore, people were equally likely to click on the correct hazard location in both the high- and low-prevalence condition. Accuracy was also fairly high at 84% (collapsed across prevalence). When simulating random clicking, click accuracy dropped to 31% (collapsed across prevalence), showing

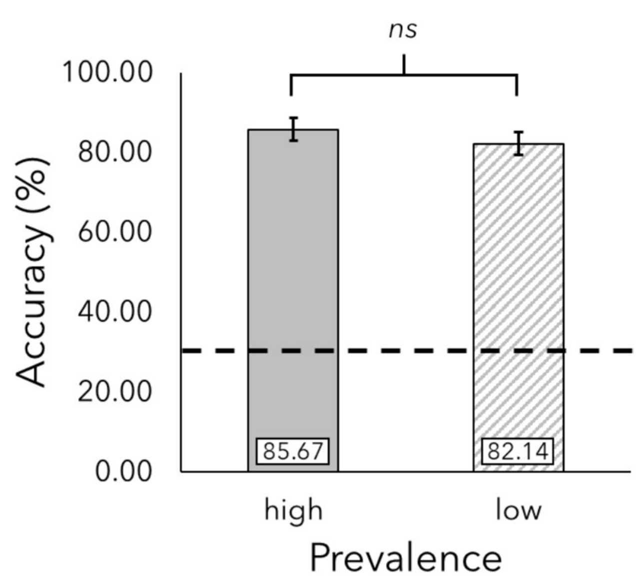

Fig. 9 No LPE for road hazard localization task in Experiment 3. There was no difference in accuracy between the high- and low-prevalence conditions in the hazard localization task. The dashed line represents what accuracy would be if participants were randomly clicking on the screen. Error bars are Morey's *SEM* (Morey, 2008)

that participants were not just randomly clicking on the screen, and had in fact correctly located the hazards.

Binary decision task

Figure 10 shows miss rate, false-alarm rate, d' and criterion. Miss rate was significantly higher in the low-prevalence condition compared with high, $t(15) = 4.02$, $p = .001$, Cohen's $d = 1$. False-alarm rate was significantly lower in the low-prevalence condition compared with high $t(15) = 5.69$, $p < .001$, Cohen's $d = 1.42$.

There was no difference in d' between the high- and low-prevalence conditions, $t(15) = 1.98$, $p = .07$, Cohen's $d = .50$. Criterion was significantly higher in the low-prevalence condition compared with high, $t(15) = 7.23$, $p < .001$, Cohen's $d = 1.81$.

Discussion

This pattern of results shows that in the low-prevalence condition, participants successfully located the hazard (with equal accuracy to the high-prevalence condition), but then said they did not need to respond to it. Therefore, any benefit of the similarity search in the localization task did *not* transfer to the binary decision task.

General discussion

Many interventions have been tested in an effort to reduce or even eliminate the LPE (e.g., Fleck & Mitroff, 2007; Hadjipanayi et al., 2023; Schwark et al., 2012). A more recent approach found that having participants identify the most target-like item (requiring a "target-present" response on each trial, eliminated the LPE (Taylor et al., 2022). However,

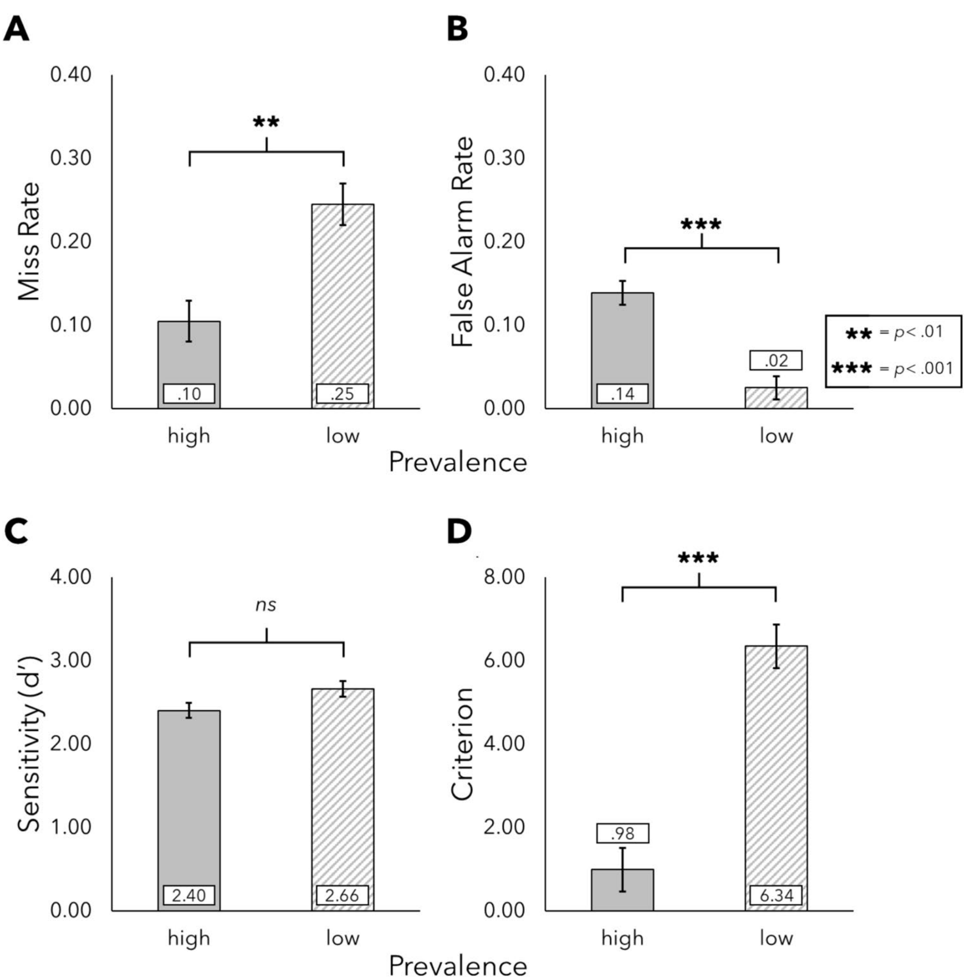

Fig. 10 LPE for the Binary Decision Task in Experiment 3. **A** Misses were significantly higher in the low-prevalence condition compared with high. **B** False alarms were significantly lower in the low-prevalence condition compared with high. **C** Sensitivity was not impacted by prevalence. **D** Criterion was significantly higher (more conservative) in the low-prevalence condition compared with high. Error bars are Morey's *SEM* (Morey, 2008)

many real-world searches require a binary decision about the selected target (e.g., "Is this item in a suitcase really a weapon?" "Do I need to hit the brakes?"). Across three experiments, we investigated if similarity search extends to binary decisions that need to be made about the target in classic search paradigms and with more naturalistic stimuli. Overall, we found that similarity search can sometimes be helpful in target identification but did not translate to binary decision-making.

Does similarity search improve target selection during low prevalence?

In Experiment 1, we found that the LPE was still present even under regular similarity search conditions. While this appears to contrast with Taylor et al.'s (2022) results, we note that any effect of prevalence is relatively small in both studies. The similarity search condition in Taylor et al.'s (2022) Experiment 3 (the most similar to our own similarity-search-only condition) showed a (nonsignificant) difference in miss rate between the high and low prevalence of 4.7%. Our observed difference for Experiment 1 (similarity search only) was 8.6% (and was significant). However, in both cases, this is notably lower than the 17.4% difference in miss rate Taylor et al. (2022) found with a conventional present/absent search. We want to emphasize that, although we observed an LPE in the similarity search tasks, it is possible that this represents a reduced LPE relative to what it would have been if participants completed a present/absent search. It is possible that similarity search attenuates, but in our case did not fully eliminate, the LPE.

Conversely, in Experiment 2, we found that there was no LPE present for the similarity-search-only task, which more closely aligns with what Taylor et al. (2022) found. The difference in accuracy between low and high prevalence was also more similar to Taylor et al. (2022), at 5.6% (not significant). There was an LPE present for the similarity search portion of the similarity search & binary decision task. For this task, feedback was only ever given about the binary decision (and not the search task), so participants could not learn anything about the accuracy of their search, and thus could not use the feedback to improve their search performance. As feedback is an important moderator of the LPE (Lyu et al., 2021; Schwark et al., 2012; Wolfe et al., 2007), we speculate that it may be the difference in feedback driving this effect.

For both Experiments 1 and 2, we found no difference in average reaction times between the high- and low-prevalence conditions on the similarity search task. This shows that participants were not terminating their search early in the low-prevalence condition. This is in line with predictions made in Taylor et al. (2022), where they speculated that similarity search increases the quitting threshold because participants must make a target present response each trial. This fits with the two-stage LPE model proposed by Wolfe and Van Wert (2010) where there is a quitting threshold ("have I found the target yet?") and a decision criterion ("Is this item the most *T*-like shape?"). We note that in similarity search, the criterion is based on a continuum of *T*-like shapes and does not require participants to classify the target as a true *T* or a non-*T*. With the inclusion of the binary decision task, we may have added a third step to this process, where the final decision criterion is "is my selected target a true *T*?"

We did not do the same set of reaction time analyses for Experiment 3, as it was not a search task (participants clicked on a grey square after watching the video). As such, response times on this task are not particularly informative for determining the underlying search processes.

Following the results of Experiment 2, we also find that similarity search may be useful in other contexts. Consistent with the findings of Taylor et al. (2022), we did not observe an LPE in Experiment 3 when participants engaged in similarity search during a road-hazard-detection task. In standard road-hazard-detection tasks (i.e., present/absent detection tasks), observers are more likely to miss the road hazards when they are rare (Kosovicheva et al., 2023). Therefore, instructing participants to find the most hazardous location in the video on every trial *did* help participants locate the real road hazards.

Inflating the target prevalence via similarity search likely helped maintain participants' expectations for a target-present response on every trial, as suggested by Taylor et al. (2022). Although criterion cannot be directly measured in the hazard localization task (there is no possible way to "false alarm" or "correctly reject" in this paradigm), we speculate that observers are widening their criteria for what constitutes a "hazardous location." In order to select a hazardous location, they need to broaden their criteria to include locations that pose absolutely zero danger to the driver (e.g., a traffic cone on the sidewalk).

Another plausible reason for why we observed a significant LPE with similarity search in Experiment 1 (and in one condition of Experiment 2), but not in in Experiment 3, is that the nature of the two tasks (*T* search and road-hazard detection) are different. Specifically, the road-hazard-detection task pertains less to "search" and is more to do with detection. Task differences are also reflected in overall performance. Average accuracy for *T* detection during high prevalence (collapsed across task type) was 57% for Experiment 1 and 60% for Experiment 2 (comparable with the average accuracy in Taylor et al., 2022). Conversely, average accuracy for road-hazard detection was 85% (collapsed across prevalence), suggesting that the road-hazard-detection task was overall easier. In other words, task difficulty may moderate the effectiveness of similarity search.

Does similarity search extend to binary decisions about the targets?

In both Experiments 1 and 2, we found that there was also an LPE for the binary decision that followed target selection. Importantly, this LPE was present when participants correctly found the target during the search task. Even when the target was successfully located, participants were less willing to label it as a true *T* during periods of low prevalence compared with high. The difference in miss rate between high and low prevalence was substantial and comparable with what is commonly seen in other LPE studies (where miss rate in the low-prevalence condition is approximately twice what it was for high prevalence; Wolfe et al., 2007). In line with previous LPE studies, this manifested as a difference in criterion, but not in sensitivity (Wolfe et al., 2007).

This double LPE is interesting, as it suggests a dissociation between target detection and the subsequent binary decision. If similarity search shifts participants' criterion to become more liberal by requiring them to identify a target on each trial, there should not have been a large LPE for the binary decision. Instead, it appears that participants adopt multiple criteria regardless of whether or not the similarity search helps them find the targets.

Though we did not find an LPE in Experiment 3 for the hazard localization task, we did find an LPE for the binary decision task. Even when participants clicked on the correct location of the hazard, they were less willing to classify it as a hazard during periods of low prevalence compared with high. The criterion for what constitutes a road hazard in the search task may have become very liberal. However, the criterion for binary response selection (e.g., braking vs. not braking) became more conservative during low prevalence.

Across the three experiments, we find that similarity search does not transfer to binary decisions that need to be made about the selected target, regardless of whether or not similarity search improves target detection. Within a signal detection framework, this suggests that observers are maintaining different criteria for the similarity search and the subsequent decision. These results are also consistent with a previous experiment by Wolfe et al. (2005), in which participants reported the presence of any one of four types of targets, each with its own prevalence (chosen from 1%, 5%, 10%, or 34%). This task raised the probability of any target being present to 50% of trials. Nevertheless, the results showed that participants maintained separate decision criteria for different items that had different prevalence rates; observers successfully found the common items, while rare items were missed.

The results of Wolfe et al. (2005) align with the current study, in which there was an inflated prevalence of targets during detection or localization ("most *T*-like shape"; "most hazardous location"). However, the prevalence of targets that needed to be classified as "true" targets in the binary decision task was much lower, and participants adopted multiple criteria accordingly. The results of our experiments suggest that similarity search may be useful in some settings. However, applying it to cases where the participant may need to make a decision about the selected target would require further work to be adopted in real-life situations.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13414-025-03084-9.

Author note Part of this work was presented at the Vision Sciences Society (VSS) Conference in 2023. The authors would like to thank the APPLY, Cant, Fukuda, and Pratt labs, with special thanks as always to Kristina Knox and the rest of the 2020 University of Toronto Psychology Graduate Cohort.

Open practices statement All three experiments were preregistered on OSF at the following links (in order of experiment: https://osf.io/rszkg, https://osf.io/kqbrm, https://osf.io/khry6). Data and analysis scripts for all three experiments are freely available on Open Science Framework: https://osf.io/83572

Authors' contribution Greer Gillies served as lead for conceptualization, data curation, formal analysis, investigation, methodology, project administration, visualization, writing—original draft, and writing–review and editing. Anna Kosovicheva served in a supporting role for conceptualization, project administration, writing—original draft, and writing–review and editing. Anna Kosovicheva was the primary contributor for funding acquisition, software, supervision, validation, and resources.

Funding This work was funded by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant to Anna Kosovicheva (RGPIN-2022-03131).

Availability of data and materials All data are available on OSF: https://osf.io/83572

Code availability All analysis scripts are available on OSF (https://osf.io/83572). Experiment code available on request.

Declarations

Conflicts of interest The authors have no conflicts of interest to report.

Ethics approval This research was approved by the Research Ethics Board at the University of Toronto (#41533). The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication Participants signed informed consent regarding publishing their data.

References

Baddeley, A. D., & Colquhoun, W. P. (1969). Signal probability and vigilance: A reappraisal of the 'signal-rate' effect. *British Journal of Psychology, 60*(2), 169–178.

Beanland, V., Lenné, M. G., & Underwood, G. (2014). Safety in numbers: Target prevalence affects the detection of vehicles during simulated driving. *Attention, Perception, & Psychophysics, 76*, 805–813.

Broadbent, D. E., & Gregory, M. (1965). Effects of noise and of signal rate upon vigilance analyzed by means of decision theory. *Human Factors, 7*(2), 155–162.

Colquhoun, W. V. (1961). The effect of 'unwanted' signals on performance in a vigilance task. *Ergonomics, 4*(1), 41–51.

Cox, P. H., Kravitz, D. J., & Mitroff, S. R. (2021). Great expectations: Minor differences in initial instructions have a major impact on visual search in the absence of feedback. *Cognitive Research: Principles and Implications, 6*, 1–14.

Evans, K. K., Birdwell, R. L., & Wolfe, J. M. (2013). If you don't find it often, you often don't find it: Why some cancers are missed in breast cancer screening. *PLOS ONE, 8*(5), e64366.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science, 18*(11), 943–947.

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 126.

Hadjipanayi, V., Ludwig, C. J., & Kent, C. (2023). Graded prioritization of targets in search: Reward diminishes the low prevalence effect. *Cognitive Research: Principles and Implications, 8*(1), 52.

Josephs, E., Fosco, C., & Oliva, A. (2024). Effects of browsing conditions and visual alert design on human susceptibility to deepfakes. *Journal of Online Trust and Safety*, *2*(2). https://doi.org/10.54501/jots.v2i2.144

Kosovicheva, A., Wolfe, J. M., & Wolfe, B. (2023). Taking prevalence effects on the road: Rare hazards are often missed. *Psychonomic Bulletin & Review, 30*(1), 212–223.

Kunar, M. A., Watson, D. G., & Taylor-Phillips, S. (2021). Double reading reduces miss errors in low prevalence search. *Journal of Experimental Psychology: Applied, 27*(1), 84.

Lyu, W., Levari, D. E., Nartker, M. S., Little, D. S., & Wolfe, J. M. (2021). Feedback moderates the effect of prevalence on perceptual decisions. *Psychonomic Bulletin & Review, 28*(6), 1906–1914.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology, 1*(1), 6–21.

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software, 3*(22), 470.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*(2), 61–64.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ..., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rich, A. N., Kunar, M. A., Van Wert, M. J., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision, 8*(15), 15–15.

Schwark, J., Sandry, J., MacDonald, J., & Dolgov, I. (2012). False feedback increases detection of low-prevalence targets in visual search. *Attention, Perception, & Psychophysics, 74*, 1583–1589.

Song, J., & Wolfe, B. (2024). Highly dangerous road hazards are not immune from the low prevalence effect. *Cognitive research: principles and implications, 9*(1), 6. https://doi.org/10.1186/s41235-024-00531-3

Taylor, J. E. T., Hilchey, M. D., Weidler, B. J., & Pratt, J. (2022). Eliminating the low-prevalence effect in visual search with a remarkably simple strategy. *Psychological Science, 33*(5), 716–724.

Thomson, D. R., Besner, D., & Smilek, D. (2016). A critical examination of the evidence for sensitivity loss in modern vigilance tasks. *Psychological Review, 123*(1), 70.

Van Wert, M. J., Horowitz, T. S., & Wolfe, J. M. (2009). Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception, & Psychophysics, 71*(3), 541–553.

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology, 20*(2), 121–124.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature, 435*(7041), 439–440.

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General, 136*(4), 623.

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of Vision, 13*(3), 33–33.

Wolfe, B., Seppelt, B., Mehler, B., Reimer, B., & Rosenholtz, R. (2020). Rapid holistic perception and evasion of road hazards. *Journal of Experimental Psychology: General, 149*(3), 490–500. https://doi.org/10.1037/xge0000665