



# Exaggerated prevalence effect with the explicit prevalence information: The description-experience gap in visual search

Hanshu Zhang<sup>1</sup> · Joseph W. Hout<sup>1</sup>

Published online: 17 June 2020  
© The Psychonomic Society, Inc. 2020

## Abstract

Despite the increasing focus on target prevalence in visual search research, few papers have thoroughly examined the effect of how target prevalence is communicated. Findings in the judgment and decision-making literature have demonstrated that people behave differently depending on whether probabilistic information is made explicit or learned through experience, hence there is potential for a similar difference when communicating prevalence in visual search. Our current research examined how visual search changes depending on whether the target prevalence information was explicitly given to observers or they learned the prevalence through experience with additional manipulations of target reward and salience. We found that when the target prevalence was low, learning prevalence from experience resulted in more target-present responses and longer search times before quitting compared to when observers were explicitly informed of the target probability. The discrepancy narrowed with increased prevalence and reversed in the high target prevalence condition. Eye-tracking results indicated that search with experience consistently resulted in longer fixation durations, with the largest difference in low-prevalence conditions. Longer search time was primarily due to observers re-visited more items. Our work addressed the importance of exploring influences brought by probability communication in future prevalence visual search studies.

**Keywords** Visual search · The description-experience gap · Prevalence effect

## Introduction

The effects of target prevalence are increasingly studied topics within visual search research (Wolfe et al. 2005, 2007; Schwark et al. 2013, 2012; Peltier & Becker 2016, 2017a, b; Evans et al. 2013; Ishibashi & Kita 2014; Ishibashi et al. 2012; Lau & Huang 2010; Rich et al. 2008; Van Wert et al. 2009). While prevalence effects on visual search have been widely discussed in some scenarios due to the practical importance of understanding search processes when targets are rare, such as medical image evaluation (Kundel 1982; 2000) and luggage screening (Wolfe et al., 2013), there are limited discussions on how target probability is understood and interpreted by observers in the unbalanced search environment. Meanwhile, research in judgment and

decision-making indicates that people treat information about event probabilities differently depending on how that probability information is communicated (Hertwig et al., 2004; Hertwig & Erev, 2009). To jointly explore the two related research areas, the goal of our current work is to systematically examine how visual search is affected by the way target prevalence information is given to observers, specifically, whether they receive explicit information about prevalence or learn the prevalence from experience. We will first give a brief overview of the prevalence effect in visual search and the description–experience gap in decision-making, then discuss our approach to investigating description–experience differences in visual search.

## Induced prevalence visual search

Compared to a balanced present/absent context, observers miss targets at much higher rates when targets are rare (e.g., from 7% miss rate to 30% miss rate; Wolfe et al. (2005)). This increase in missed targets with a decrease in target prevalence is referred to as the *prevalence effect* (see Horowitz (2017), for a review). These general patterns have been found with both ways of communicating

✉ Hanshu Zhang  
zhang.180@wright.edu

Joseph W. Hout  
joseph.hout@wright.edu

<sup>1</sup> Department of Psychology, Wright State University,  
Dayton, OH, USA

target prevalence information, learning the prevalence from experience and explicit information.

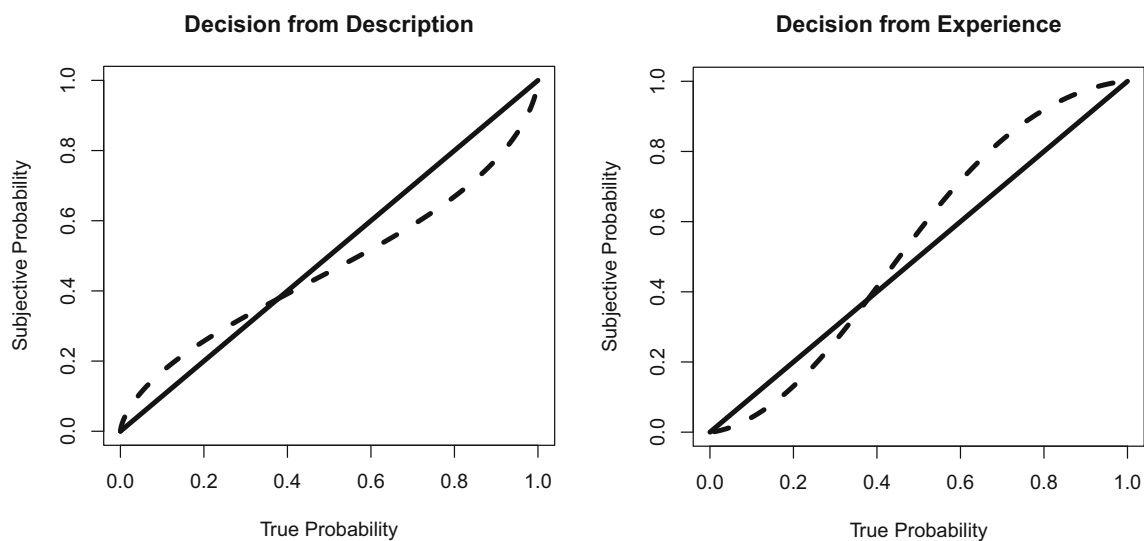
In experimental settings, experience-based search can either be based on learning during the main blocks, i.e., observers start searching without any prevalence information (e.g., Wolfe and Van Wert (2010), Experiment 2) or observers experience the target prevalence in practice trials prior to the main data collection blocks (e.g., Peltier and Becker (2016)). Experience-based prevalence knowledge provides a clear analog to applied settings, and can offer insight into those domains. For example, Evans et al. (2013) demonstrated that radiologists have relatively higher miss rates in contexts in which they rarely observe cancer compared to contexts in which cancer is more commonly observed. In a more extreme case, the “experience” effect had influence on search performance even when the “actual” prevalence is fixed; even when there were no targets at all, observers still followed false feedback reporting more frequently that target was present (Schwark et al., 2013).

In contrast to “experience”-based search, in some research observers are given vague prevalence information, e.g., that targets are “rare” or appear “often” (e.g., Wolfe et al. (2005)). The evidence of the influence of this explicit prevalence information visual search performance is more divided. For example, some research found that trial-by-trial indicators of whether targets were likely or not did *not* significantly moderate the prevalence effect; instead, observers relied more on their accumulated experience of target prevalence (Lau & Huang, 2010; Ishibashi et al., 2012). An additional study reported that the “incorrect” number did not affect radiologists’ accuracy or efficacy on reporting abnormal pulmonary nodular lesions (Reed et al., 2011). Even though the influence of explicit prevalence information on overall search accuracy was dominated

by experience, the scrutiny time and the number of fixations increased with cues, indicating higher prevalence. Moreover, the explicit information influence was more apparent for target-absent images (Reed et al., 2014). Therefore, descriptive cues were less effective in instructing observers’ search outcomes, but potentially affect the search process, which can be driven by expectations. Nocum et al. (2013) tested “naïve” observers on various prevalence expectations and argued that sensitivity and fixation time were changed due to the expectations. Recent studies that framed the clinical information of images into three different prevalence expectations context found higher expectation for abnormal reading increased search and dwell time and resulted in more false-positive reports (Littlefair et al. 2016, 2017).

### The description-experience gap

While evidence of influence of explicit information on search outcomes is varied, whether or not probabilities are communicated explicitly has a clear influence on risky decision-making. This difference between choice based on explicit probability instructions compared to learned probability information is usually couched within *prospect theory* (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992), and particularly the implied mapping between objective and subjective probabilities. Prospect theory posits that decision-makers subjectively *overweight* small probabilities and *underweight* moderate and high probabilities (Fig. 1, left panel). This pattern allows prospect theory to explain, among other things, choices in extremely low probability events such as purchasing lottery tickets, gambling, and insurance. In contrast, Hertwig et al. (2004) demonstrated that decisions based on experienced



**Fig. 1** An illustration of subjectively weighted probability in the scenario of decision from description (*left*) and decision from experience (*right*)

probabilities, rather than explicit probabilistic information, followed a pattern opposite from prospect theory (Fig. 1, right panel). In decisions from experience, observers behave as if they *underweight* small probabilities and *overweight* large probabilities.

Recently, Wulff et al. (2018) reviewed the potential determinants of the description-experience gap in the decision-making literature: Decision-makers may rely on small samples, weight more on the recent observations (i.e., *recency*), or subjectively reverse the probability weighting. However, there was no single determinate that can explain the “gap” (also see Hertwig and Erev (2009)). Though there is a lack of evidence in the previous visual search studies showing the difference between the search with prevalence-learned experience and from description, the core prevalence effect parallels the pattern of choices from experience—rare events are treated as though they were even more rare (higher miss rates). The match between “prevalence effect” observation and “decision from experience” motivates the systematical investigation of whether the same “gap” exists in the prevalence visual search as well. Second, we examine the subjective interpretation of target prevalence—whether observers relatively overweight or underweight target probability depending on whether they are given explicit prevalence information or infer prevalence from experience. For example, if observers’ behavior follows the pattern of behavior in risky choice, then when they are told the target is highly prevalent, they will act as though it is slightly less prevalent, and when they must learn the same prevalence from experience they will act as though it is more prevalent (even higher).

## Current research

To summarize, our main goal in the present research was determining whether the “description-experience gap” is reflected in prevalence visual search performance. We hypothesized that observers would treat probabilistic information in accordance with prospect theory when given explicit probability information. Thus, we predicted observers would treat very low prevalence contexts as if the target probability were higher and treat moderate and high prevalence contexts as though the target probability was lower. Second, we hypothesized that when observers searched based on their accumulated experience, the same mechanisms that lead to the description-experience gap in decision-making would influence their search performance. We predicted observers would underweight small target probabilities and overweight moderate and high target probabilities.

## Theoretical analysis and prediction

In both the description conditions and experience conditions, the true prevalence is determined, but it is not possible to directly measure observers’ subjective prevalence. In the decision-making literature, subjective probabilities are estimated based on a modification of a normative model of choice, i.e., if  $u_i$  are the values of possible outcomes if  $A$  is chosen, each of which would occur with subjective probability  $\pi_i$ , then,

$$\Pr\{\text{Choose } A\} \propto \sum_i u_i \pi_i.$$

We follow a similar approach by assuming that observers’ search performance can be described by a signal detection model with an optimal criterion for their subjective utility (cf. Wolfe and Van Wert (2010) and Wickens (2002)). A primary advantage of using the signal detection model is that it redescribes patterns of accuracy in terms of target discriminability and response bias. Briefly, the signal detection model assumes that a stimulus induces a random amount of evidence for the presence of a target, where there is more information on average in favor of a target when the target is present. In the model, an observer responds “target-present” whenever the evidence exceeds a criterion and absent whenever the evidence is below the criterion. When the variation in the amount of information induced by a stimulus is high enough, it is possible that evidence will exceed the criterion when no target is present and the observer will make a false-alarm error, or possible that the evidence will be below criterion when the target is present and the observer will have a miss error. The difference in the average amount of information indicating target presence when it is actually present, and the amount of information when the target is absent represents the overall target discriminability, and is labeled  $d'$ . The criterion represents the response bias; a lower criterion implies relatively more target-present responses, while a higher criterion corresponds to more target-absent responses. There are different ways to parameterize the response bias. For correspondence with previous prevalence visual search studies (e.g., Wolfe and Van Wert (2010) and Godwin et al. (2015)), we use  $c$ . Henceforth, unless otherwise specified, we are referring to  $c$  whenever we refer to *criterion* within the signal detection model. More details are included in Appendix A.

The relationship between the observers subjective interpretation of the prevalence,  $\pi$ , subjective values of

correct rejections,  $u_{CR}$ , false alarms,  $u_{FA}$ , hits,  $u_H$ , misses  $u_M$ , and the optimal criterion  $c$  is given by,

$$c_{\text{optimal}} = \frac{1}{d'} \left[ \log \frac{u_{CR} - u_{FA}}{u_H - u_M} - \text{logit } \pi \right]. \quad (1)$$

Equation 1 demonstrates that an optimal observer would change its criterion based on changes in perceived prevalence  $\pi$ . As  $\pi$  decreases,  $\text{logit } \pi$  decreases and hence the optimal criterion increases, resulting in relatively more misses. We expect the difference between prevalence from experience and prevalence from description to selectively influence  $\pi$ .

Equation 1 also indicates an effect of the response rewards and costs (utilities). Assuming that the subjective target prevalence  $\pi$  is fixed, if  $u_{CR} - u_{FA} < u_H - u_M$ , i.e., target-absent responses are generally less valuable than target-present responses, then the optimal criterion would be lower, resulting in relatively fewer misses. In contrast, if the values  $u$ . for correct and incorrect outcomes are balanced:  $u_{CR} - u_{FA} = u_H - u_M$ ; then the criteria for reporting targets would be entirely driven by the (subjective) target prevalence and discriminability. Researchers have found effects of the relative response values on the prevalence effect. For example, Navalpakkam et al. (2009) found that with high penalty on misses of reporting targets, observers were able to report more “target-present” when the target prevalence was low and reached the optimal performance. Thus, in addition to the target prevalence and how it is communicated, we also manipulate the value,  $u$ ., associated with each response.

The use of signal detection theory in studying prevalence visual search has some precedent. In particular, it is one of two main components of the dual threshold model from Wolfe and Van Wert (2010). Following the dual threshold model (Wolfe & Van Wert, 2010), we extend the signal detection model to include a random accumulation process for quantifying the effects of varying target prevalence and reward schemes on search strategy. When target prevalence is low, observers tend to end their search sooner and thus have fast target-absent responses. This effect has been demonstrated in both behavioral studies (Lau & Huang, 2010; Ishibashi & Kita, 2014; Ishibashi et al., 2012) and eye movements studies (Peltier & Becker, 2016; Godwin et al., 2015). The dual threshold models this phenomenon as an effect of a higher threshold for the random accumulation process to reach for terminating. In addition to the effect of prevalence on shifted criteria  $c$ , we expect the quitting threshold to increase with higher penalties for misses through the manipulation of reward scheme  $u$ . as observed in Navalpakkam et al. (2009).

We expect that observers’ subjective weighting of target probability thus can be observed through the measurement of criteria shift and quitting threshold, as described in the dual threshold model. As Fig. 2 sketches, we hypothesized interaction between information manipulation and prevalence on criteria (Fig. 2a) and quitting threshold (Fig. 2c)—e.g., observers search in low prevalence with explicit experience would have faster and more “target-absent” responses (the black line on the left end). Second, the penalty on the missed target reports would ameliorate the prevalence effect (Fig. 2b and d). We will elaborate on our manipulation implementations in the following section.

## Methods

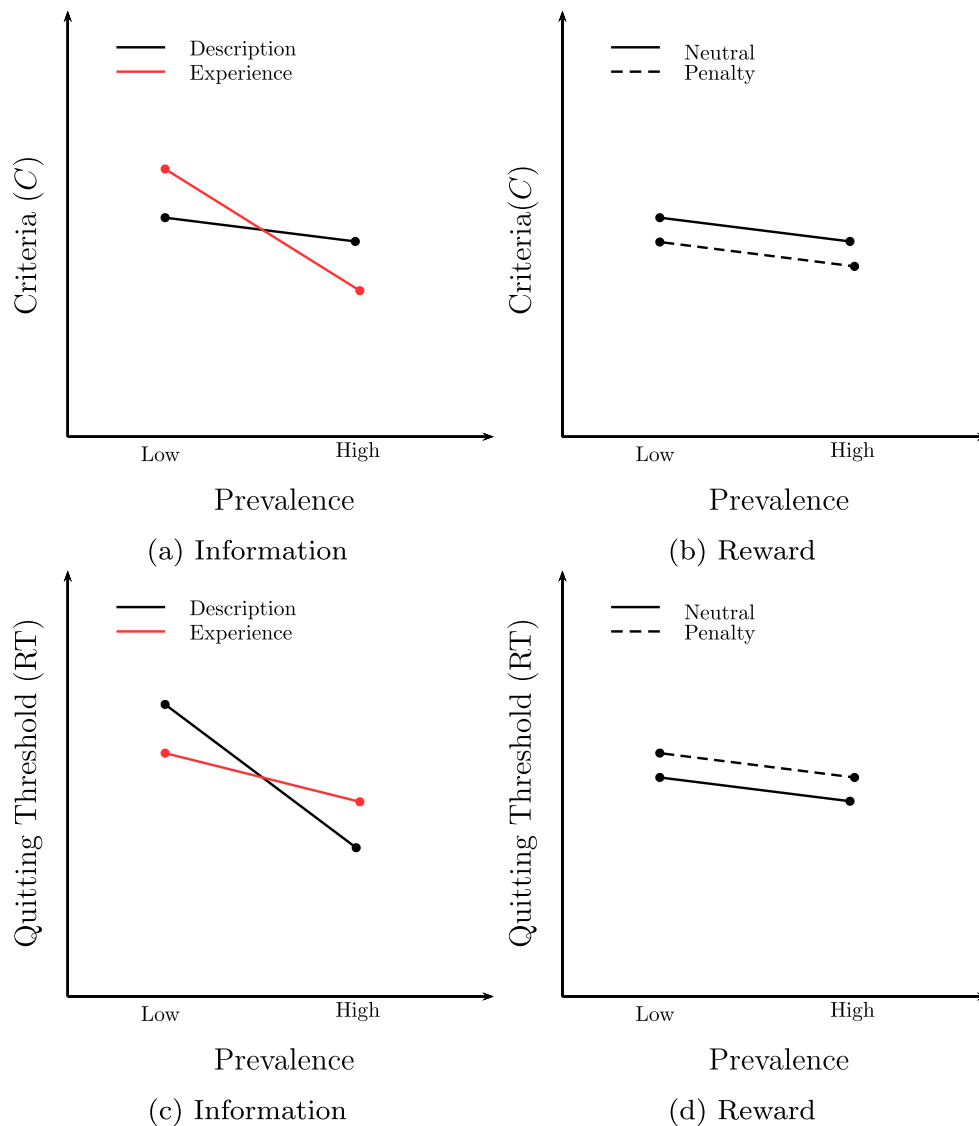
### Participants

We targeted 20 observers for analysis in this study whose performance is above chance—50% overall accuracy averaged across all prevalence levels—based on sample sizes in Wolfe and Van Wert (2010) and the intention of examining individual participant level performance of small number of group over large number of observations (cf. Smith and Little (2018)). Ultimately, we collected data from 22 observers (age: 18–59; female: 13). Two observers whose performance was lower than 50% accuracy and were excluded from further analyses. All observers reported normal or corrected to normal visual acuity and no difficulty understanding English. Observers were reimbursed \$40 for finishing the entire task and were motivated by receiving \$20 as an extra bonus for being in the top 15% for search performance based on points they achieved defined by the overall reward schemes.

### Experimental design

The study consisted of four fully crossed design sessions, combining two different search conditions (i.e., experience/description) and two different reward schemes (i.e., penalty/neutral; see Table 1). Each session was expected to last about 1 h 15 min including breaks during the experiment, and observers were expected to complete no more than one session per day and scheduled to finish all four sessions within a week.

Each session consisted of four blocks with target prevalence at 0.1, 0.35, 0.65, and 0.9 in random order. Each block included 80 trials. Given that observers may be sensitive for short-term runs in prevalence (e.g., Fox and Hadar (2006)), we used a fixed random distribution of target-present trials within a block. That is, the sequence of “target-present” trials are same for all different information by session conditions—e.g., for an observer, “trial 2” always



**Fig. 2** Experimental predictions on hypothetical analysis

includes a target in 0.9 condition for both *experience* and *description* conditions.

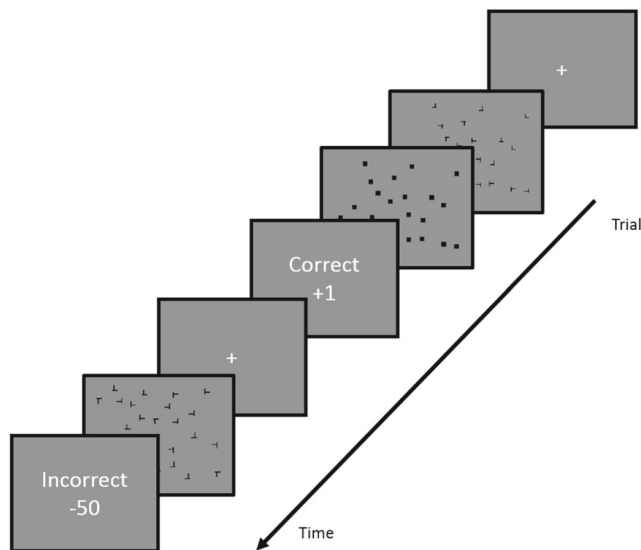
Although target/distractor discriminability is not the primary focus of our research, we also included this manipulation to verify that changes in our interpretation of variation in subjective probability was not due to variation in  $d'$  (cf. Eq. 1). To manipulate discriminability, we used two levels of distractors defined by how close they were to the targets. Rather than having homogeneous distractors, the discriminability level for a trial was determined by the proportion of distractors that were more and less similar to the target in the search field. In a high salience trial, the high perceptually discriminable distractors were twice as likely to show up as the low perceptually discriminable distractors, and vice versa for a low salience trial.

In summary, our current design deployed four different manipulations to influence search performance: *information* (description/experience), *prevalence* (high/low), *reward* (penalty/neutral), and *salience* (high/low). We observe the manipulations through dependent variables of behavior responses and eye movements.

Figure 3 depicts the basic trial structure. Each trial began with a cross in the middle of the screen. The stimulus was

**Table 1** Reward schema adapted from Navalpakkam et al. (2009)

	CR	FA	Miss	Hit
Neutral	+1	-50	-50	+1
Penalty	+1	-50	-900	+100



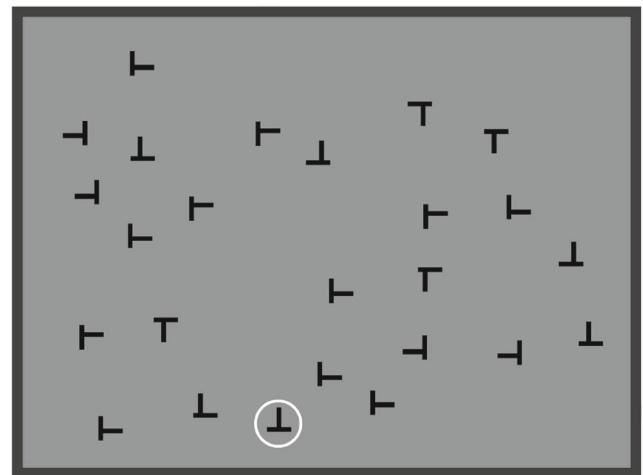
**Fig. 3** Structure of a single trial. See text for details

presented on the screen and observers were asked to respond by clicking the mouse to indicate if the target was present or not. If observers responded that the target was present, they were then asked to verify the location of the target by clicking on the item position. If observers responded that target was absent, then the trial ended. Observers received feedback and reward outcomes in points after each trial. In particular, observers were given details of their incorrect responses: responding target present in a target-absent trial or identifying the wrong item in a target-present trial (i.e., false alarm),<sup>1</sup> or responding target-absent in a target-present trial (i.e., miss). Observers were instructed to respond as fast and accurately as possible and were told that the trial would end after 15 s. To encourage timely responses, not responding within the 15 s resulted in a high penalty—twice the “miss” points.

## Materials

The experiment was programmed using PsychoPy (Peirce, 2007). Stimuli were a field of T-shaped items randomly placed on the screen. An example search field is shown in Fig. 4. Each search array consisted of 25 items. Each item subtended  $1^\circ \times 1^\circ$  visual angle (VA). The perceptual discriminability between targets and distractors was controlled by distance of the crossbar away from the center (offset  $0.08\text{--}0.2^\circ\text{VA}$ ). The larger the VA, the more dissimilar the item was from the target and more

<sup>1</sup>The reward schemes took responses of identifying a wrong item in the target present trials as false alarms. We also analyzed the case when the responses were considered as hits (consistent with previous studies where observers did not need to verify the target position) or misses (failed to report a target), and there were no qualitative changes relevant to our conclusion.



**Fig. 4** Stimuli were a field of T-shaped items randomly placed on the screen. The example given was a target present trial in which the target was circled for displaying purpose. The trial was also featured as low salience that more low salience discriminability distractors were presented

easily being identified as a distractor. High perceptually discriminable distractors (offset  $0.125 - 0.2^\circ\text{VA}$ ) and low perceptually discriminable distractors (offset  $0.08 - 0.11^\circ\text{VA}$ ) were sampled to compose high/low salience search scenario for each trial. The stimuli were presented on a 20” Sony Trinitron CRT monitor with a resolution of  $1600 \times 1200$  pixels and a refresh rate of 85 Hz. Observers viewed the screen at a distance of 90 cm.

Eye movements were tracked using an EyeLink 1000 eye-tracker at a 500-Hz sampling rate and only the right eye was recorded. Observers were required to use a chin rest to stabilize head position and asked to move their heads as little as possible.

## Procedure

Information condition (description/experience) and reward (penalty/neutral) were fully crossed and each combination was administered on a different day. All subjects followed a pattern experience–description–experience–description for the information condition. Being given explicit prevalence information in the *description* condition first might offer observers an expectation of prevalence in the subsequent *experience* condition in which they were supposed to accumulate target prevalence information.<sup>2</sup> The reward condition was counterbalanced across participants.

<sup>2</sup>Though we did not counterbalance all order effects, we limited participants to one session per day to mitigate carry-over effects. Our further test on learning effects and effects of exposure to the description condition indicated there was no discernable difference between the first and second experience session at any level of prevalence.

Before the experiment started, the eye-tracker was calibrated using a nine-point calibration routine. After the calibration, observers were given the instruction about the task. Observers were instructed to find the target symmetrical T among various asymmetrical distractors T's as accurately and quickly as possible. Additionally, the instruction also gave observers an illustration of all possible distractors showing that some distractors were more difficult to identify.

In the *experience* session, the experimenter told the participant that there would be four blocks and some blocks included more targets than other blocks. The experiment then began with a block indicator—i.e., “Block 1”. In the *description* sessions, the experimenter told the participant that there would be four blocks, but the probability of target was display with the block indicator—i.e., “Block 1 0.35”. Though not explicitly described in the experimental instruction, observers were given a verbal example to help them understand the interpretation of target prevalence—e.g., “35% means there are 35 trials that include a target in every 100 trials on average”.

Each block lasted 80 trials. To maintain consistent starting points for the visual search and to check for drift in the eye-tracking measurements, each trial started only after the observer fixated a cross ( $2.5^\circ$  VA of center) at the center of the screen (Fig. 3). After each block, observers were required to take a break and then performed the calibration routine again before returning to the task for next block.

## Bayesian analyses

We report our results in terms of the posterior distribution of the appropriate model (signal detection or generalized linear model) parameters and in terms of Bayes factors (BF). When the models were available in the BayesFactor package (Morey & Rouder, 2018), we used it for both posterior estimation and Bayes factor estimation. For models that are not included in that package, we implemented the model in Stan (Stan Development Team, 2018) to estimate posterior distributions and fed results into the bridgesampling package (Gronau & Singmann, 2018) to estimate Bayes factors.

We used BF in a manner similar to the Bayesian ANOVA tests implemented in R (Morey & Rouder, 2018). The Bayes factor indicates the relative support likelihood of the observed data under two alternative models. For example, if  $BF = 4$ , the observed data are four times more likely to have occurred under one model relative to the other model. For qualitative interpretation of the Bayes factor scale, we use (Jeffreys, 1961). Thus, we would state that there is moderate evidence for that model if  $BF = 4$  (see also Lee and Wagenmakers (2014)). This approach reframes the traditional null-hypothesis testing question of whether a

factor is significantly different from zero as a question of whether the data are more likely under a model that includes the factor. For example, rather than testing if the variance across levels of an interaction is higher than within, our analysis compares how likely the data are under a model with the interaction to a model with only the main effects. In analyses with many different possible models (e.g., all subsets of main effects and interactions), we report only the top models, i.e., those models that have the highest relative likelihood among the possible models. This allows us to focus on discussing the factors and interactions that are most likely given the observed data.

The posterior distribution describes the estimation of the relative plausibility of different parameter values, conditional on our observed data and model (cf. McElreath (2018)). For each posterior distribution, we present the 95% credible interval that was a range of posterior distribution, indicating that the interval was 95% likely to contain the true value of the parameter (Lee, 2018).

## Results

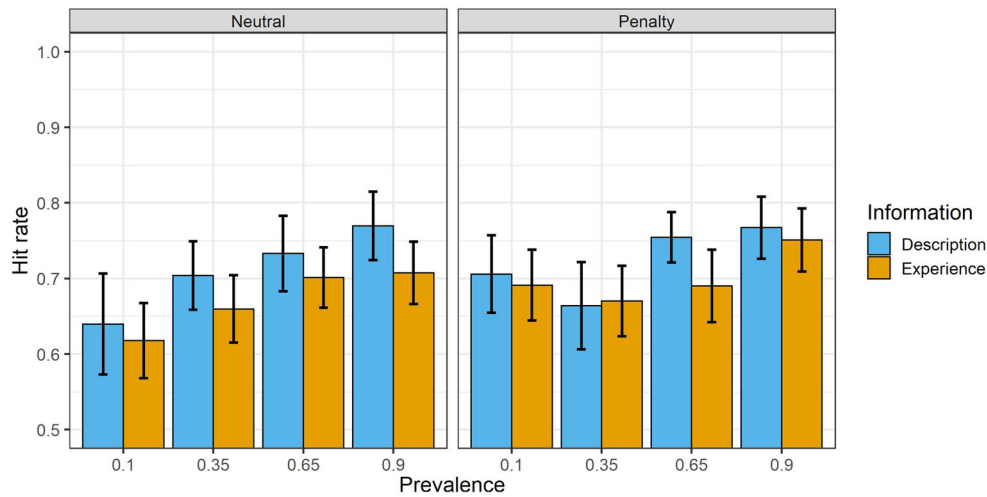
We collected observers, accuracy, response time, and eye-movement trajectories for each trial. We first verify that the standard prevalence effect was replicated, then present analyses for each of the three data types. In addition to the simple effect of the prevalence, we are interested in the interaction with the two different information conditions: description and experience. To measure this difference, we focus on shifts in signal detection criteria and response times (dual threshold model; Wolfe and Van Wert (2010)), item fixation during (Peltier & Becker, 2016), and number of fixated and re-fixated items (Godwin et al., 2015).

## Accuracy analysis

Among the 20 observers who reached the accuracy criteria (i.e., 50% accuracy and better), three additional observers were excluded from the analysis due to their failure to discriminate targets from distractors for any given block (i.e.,  $d' \leq 0$ ), which left 17 observers in total in our behavior analyses. We analyzed the remaining 17 observers using a Bayesian hierarchical signal detection model, based on Rouder and Lu (2005), and implemented in Stan (Stan Development Team, 2018).

Figure 5 gives a general description of observers' accuracy when the target was present. As Fig. 5 indicates, we find the same pattern as previous studies reporting “prevalence effect”—as the target prevalence decreases from high to low, observers missed more targets.

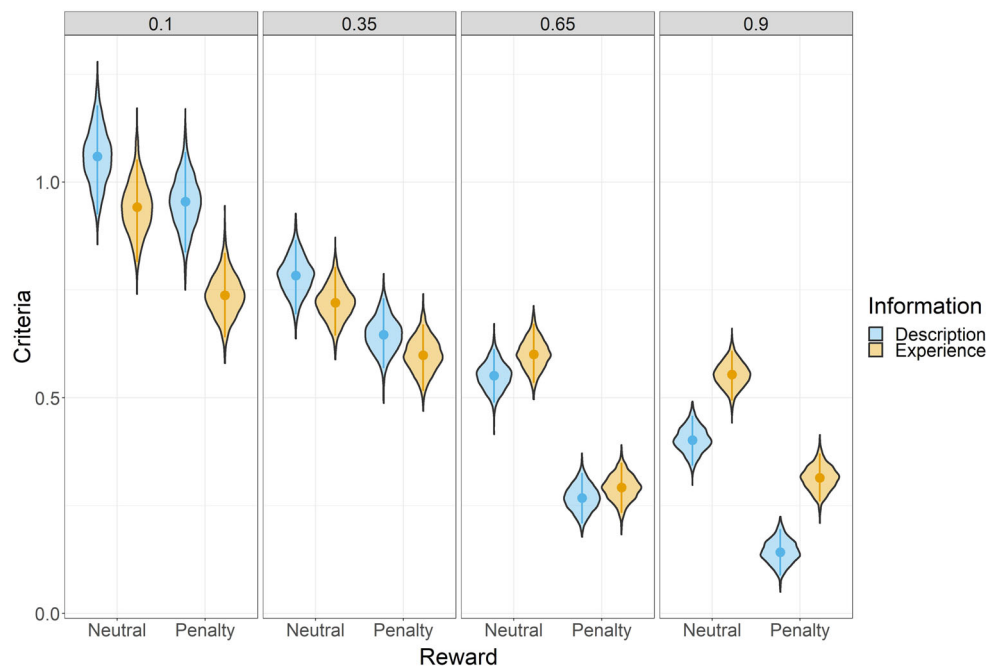
Given that the basic effect was replicated, next we applied the Bayesian signal detection analysis. For a



**Fig. 5** The accuracy when the target was present across over the different prevalence, reward, and information conditions. The *error bar* indicates standard error within each group

baseline of comparison, we estimated a model that assumed only main effects of information, prevalence, reward, and an individual subject variance factor. We then evaluated Bayes factors for a model that included all two- and three-way interactions among prevalence, information, and reward for estimating posterior distributions and evaluating evidence for our hypotheses. Compared to the baseline model, there was extreme evidence in favor of the full model that included all main factors and interactions ( $BF > 1.45 \times 10^{23}$ ).

Figure 6 shows violin plots of the posterior distribution of group-level criteria across reward levels and information levels for the best model. The 95% high-density intervals (HDI; cf. Meredith and Kruschke (2018)) are indicated with a line and the full distribution is indicated with filled shapes. The main effect of reward is clear evidence in the posterior distribution. Overall, as prevalence increased, the criterion decreased, indicating participants became relatively more biased toward responding target-present. There was also an evident effect of the reward manipulation: Criteria were



**Fig. 6** Violin plot depiction of the posterior distribution of the group-level criteria parameter from the hierarchical signal detection model over the different prevalence, reward, and information conditions



higher in the neutral condition than in the penalty condition, indicating that participants were more biased toward target-present responses in the penalty condition.

The interaction between prevalence and information condition is indicative of a description-experience gap. Figure 6 shows that criteria tended to be more biased toward target-present in the experience condition for low prevalence. This discrepancy diminished when the target prevalence increased to moderately low ( $s = 0.35$ ) and moderately high ( $s = 0.65$ ) prevalence conditions. In the high-prevalence condition, the effect of the “gap” was reversed: the criteria in the description were more biased toward target-absent responses.

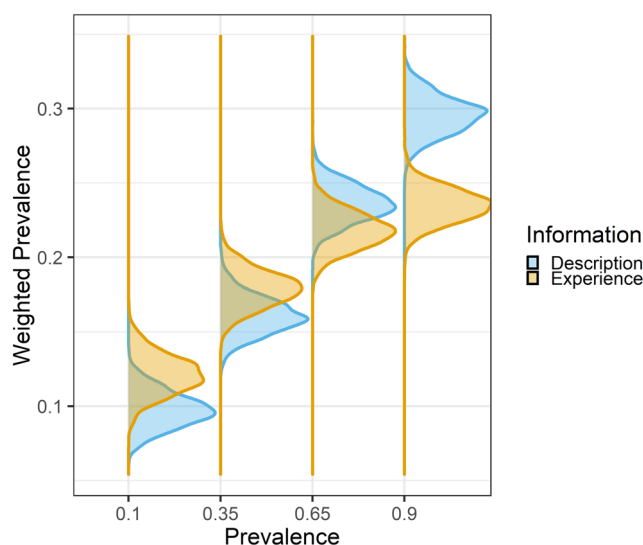
In decision-making literature, the description-experience gap is framed in terms of the mapping between true probabilities and subjective probabilities (i.e., Fig. 1). To highlight the relationship between our results of subjective probability weighting and true target probability, we used the posterior estimates of the SDT model to predict the subjective prevalence used by participants in our study. This translation rests on two assumptions: First, participants set their criterion optimally given their subjective utility; and second, they used the true point value as the utility.<sup>3</sup> More succinctly, we calculated subjective probability  $\pi$  in Eq. 1 assuming that observers performed on optimal criteria  $c$ .

Figure 7 shows the posterior group-level subjective probabilities. Contrary to the weighted probability illustration in decision-making literature, as described in Fig. 1 that observers’ performance deviated from the true target probability by either subjectively overweighting or underweighting probability, our results indicated that observers underweighted the target probabilities across all sessions.<sup>4</sup> Nonetheless, the cross-over interaction between prevalence and information condition is again evident in these posterior distributions: In the low prevalence condition, search based on experience resulted in observers weighted probability larger compared to when they were given explicit information about the probability; in the high prevalence condition, the description-guided search led to observers have larger weighted probability than in the experience-based search.

**Expectation and experience** In the experience conditions, participants had less experience when responding to trials early in a block. Similarly, in the description condition, the participants had the opportunity to gain a lot of experience

<sup>3</sup>These assumptions are almost certainly false, but does not have an effect on the qualitative relationship between subjective probability and information condition across prevalence levels and hence is less important for our purposes. In future research, we plan to add conditions that will allow for estimating subjective utility as well subjective prevalence. See Appendix A for further discussion.

<sup>4</sup>Note that this consistent underweighting may be due to our simplifying assumption about the subjective utility mapping.



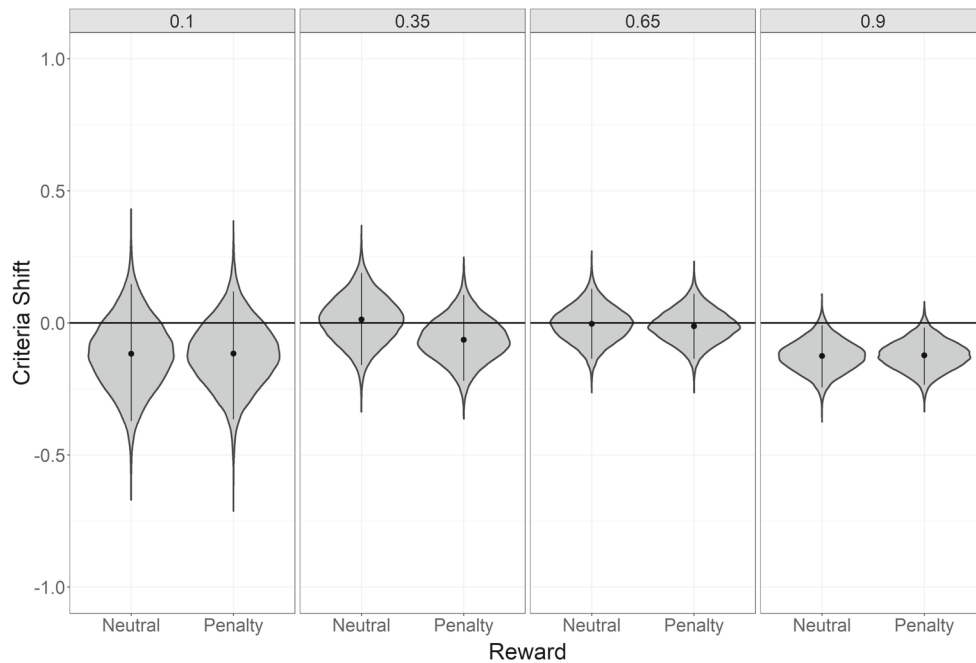
**Fig. 7** Implied probability weighting assuming optimal criteria (neutral reward condition); the *lines* indicate the median of the distributions

of the prevalence by the later trials in a block. Thus, to examine whether the effects hold when comparing the trials with the most experience in the experience condition and least experience in the description condition, we ran the same signal detection analysis applied to the first half of the trials from each of the description blocks and the latter half of trials from each of the experience blocks. The analysis indicated extremely strong evidence in favor of the full model that included all main factors and interactions ( $BF = 1.91 \times 10^{18}$ ).

Figure 8 demonstrates the evidence for differences in bias between the early trials in the description block and the late trials in the experience block with the posterior distribution of the criterion differences. The posterior difference is not clearly different from zero at low and moderate prevalence levels (Fig. 8 left three columns), indicating that the difference is less likely between 2nd half the experience blocks and 1st half of description blocks. Therefore, in these prevalence search blocks, learning from “experience” through trials resulted in similar search behavior as explicitly given the target probability information. However, the difference is still likely (HDIs below zero) in the high prevalence levels. The comparison between expectation and experience indicated that even with more experience where the target is present, informing observers that the target probability is 90% produced more “target-present” reports.

## Response times

Based on Wolfe and Van Wert (2010), we expected observers would search longer in higher prevalence conditions. Therefore, observers would have higher response



**Fig. 8** Violin plot depiction of the posterior distribution of the difference (gap) between the criterion for trials in the second half of the experience blocks (most experience) and first half of the description condition (least experience)

times reporting “target-present” with higher target prevalence, regardless of whether the target was present or not. In accordance with this expectation, we examined the effects of information, reward, prevalence, and distractor salience on target-absent response times, regardless of whether the target was present, using a Bayesian ANOVA (Morey & Rouder, 2018).

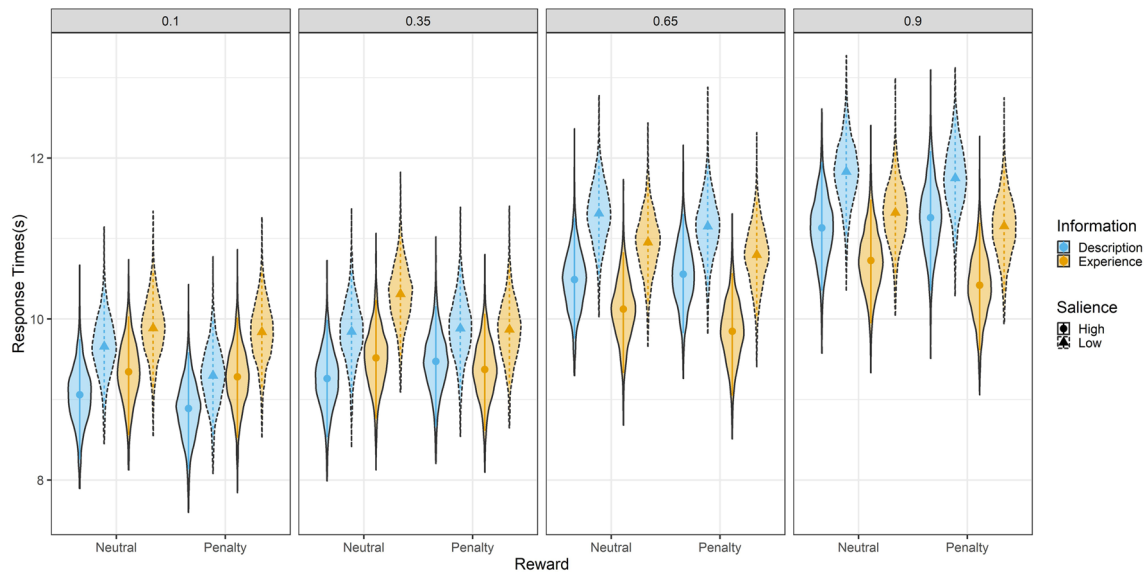
Compared to the baseline model that only included variability across subjects’ as a factor, our results indicated that the model with the highest BF included the main effects of information, reward, prevalence, salience manipulation, and an interaction between information and prevalence ( $BF > 7.62 \times 10^{324}$ ). To test how strong the observed data evidence supporting the best model in favor of the other models that also covered these factors, we compared the best model to the second and third best model. The comparisons indicated that the best model was only slightly better than the next best model, which excluded the effects of information and the interaction between information and prevalence ( $BF = 2.06$ ), and moderately better than the third best model, which added the interaction between reward and salience ( $BF = 4.16$ ).

Figure 9 gives the posterior means and HDIs from the full model including all factors and interactions. As expected from the prevalence effect: response times were higher in higher prevalence blocks—the response time increased from the left panel to the right panel. Figure 9 also reveals the effect of the salience manipulation that

more discriminable distractors led to lower response times (the side-by-side violin distribution comparison in the same color), confirming our assumption that these trials were easier. In addition, there was only anecdotal evidence supporting the effect of information and the interaction between information and prevalence when comparing the best and second-best model, conditional on our observation. We found that if the prevalence-information interaction in RT exists, it falls into the same pattern of criteria as described in the previous section: when the target was in lower prevalence, observers had lower response times in the experience conditions; in the higher target prevalence scenario, observers had higher response times in the description condition.

### Eye movement

To further understand the visual search process, we examined fixation duration and patterns of the 17 observers who had positive  $d'$ . We explored fixation duration on trials where observers correctly responded target-present (i.e., hit; Peltier and Becker (2016)). Moreover, to be consistent with our SDT analysis and quitting threshold as discussed in Wolfe and Van Wert (2010), we included all trials when observers responded target-absent for analysis of number of fixated items—that is both correct rejection and miss rather than only correct rejection as Godwin et al. (2015). Only fixations between the onset of the search array and the



**Fig. 9** The posterior distribution for group-level target-absent response times

response were used in these analyses. Fixations that were shorter than 50 ms were excluded from the data analysis (Rich et al., 2008).

A fixation was classified as being on a particular element in the display if the fixation fell within  $120 \times 120$  pixel (about  $2^\circ \times 2^\circ$  visual angle) of the center of the item. Trials were marked as invalid if the proportion of invalid samples from the eye-tracker exceeded 50%. For example, observers had too many or too long eye blinks, or the eye-tracker failed during the trial. Two additional observers were excluded from the eye movement data analysis for having an excessive number of invalid trials (i.e., > 40 trials) in a block. For the remaining 15 observers, from about 0.5% of trials were removed from further analysis based on this criterion.

**Item fixation duration** We predicted that in the high prevalence condition, observers fixated shorter on targets and longer on distractors compared to the low prevalence conditions. In addition, we assumed that the “gap” we observed in the behavioral data analysis had an effect on the fixation time as well—overweighting target probability would lead observers to fixate on the item longer.

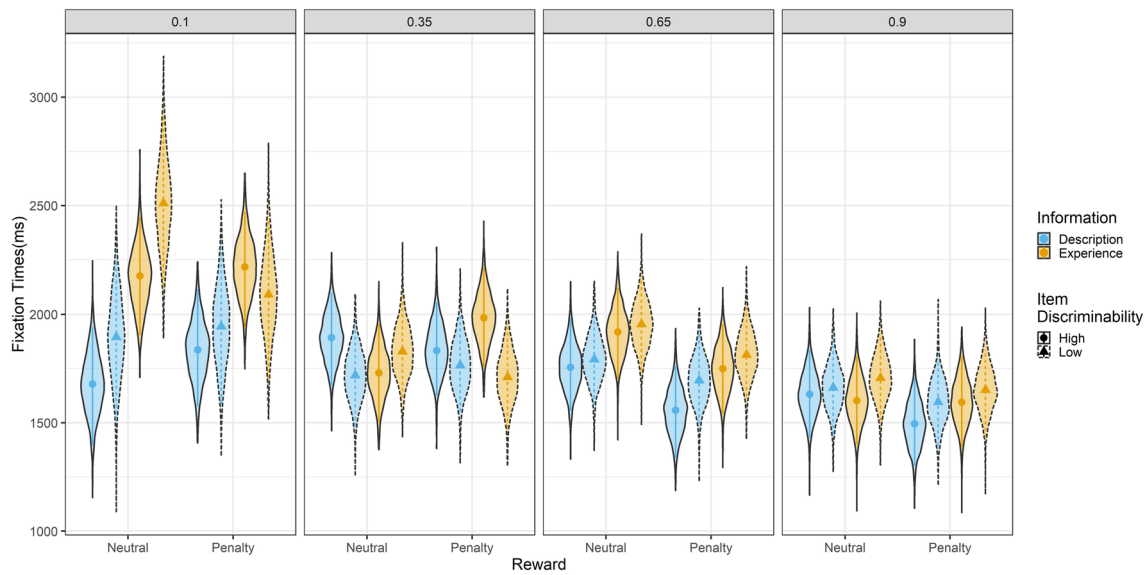
For both fixation duration on targets and distractors, we used the Bayesian ANOVA (Morey & Rouder, 2018) to compare the full model to our baseline model that only included subjects’ variance. For targets fixation duration, the most likely model included all main effects of information, reward, prevalence, salience, and the interactions between prevalence and reward, and prevalence and salience ( $BF > 7.73 \times 10^{45}$ ). The most likely model was favored over the next best model ( $BF = 4.47$ ) and third best model ( $BF = 5.62$ ) with moderate evidence, which did not

included both salience and its interaction with prevalence, or only the interaction, respectively. For distractors fixation duration, the most likely model included the main effects of information, prevalence, salience, and the interaction between prevalence and salience ( $BF > 5.97 \times 10^{28}$ ). It was favored over the next model with moderate evidence ( $BF = 3.75$ ) that did not include the information effect, and stronger evidence ( $BF = 25.50$ ) with the additional interaction between salience and information.

Figures 10 and 11 describe the posterior means and HDIs from the full model, including all factors and interactions for observed item fixation duration. Consistent with previous findings on the prevalence effect, the Bayesian analysis indicated that there was a main effect of prevalence on the fixation duration of both targets and distractors. Additionally, the results also indicated the effect of reward on target fixation times and the effect of information on distractors fixation times. Contrary to our expectations, we failed to find evidence showing that there was an interaction between the prevalence and information.

**Visited/re-visited items** To further examine the nature of the changes in search duration, we analyzed the total number of fixated items and the number of fixations on previously fixated items. Recall that we predicted that increases in search time were driven by an increase in the number of fixated and re-fixated items. Specifically, a re-visitation is counted when an item is fixated again after observer fixating on another item.

We analyze the number of fixations across target present and target absent trials when observers quit the search, i.e., responded target-absent, using Bayesian Poisson regression implemented in the R package BRMS

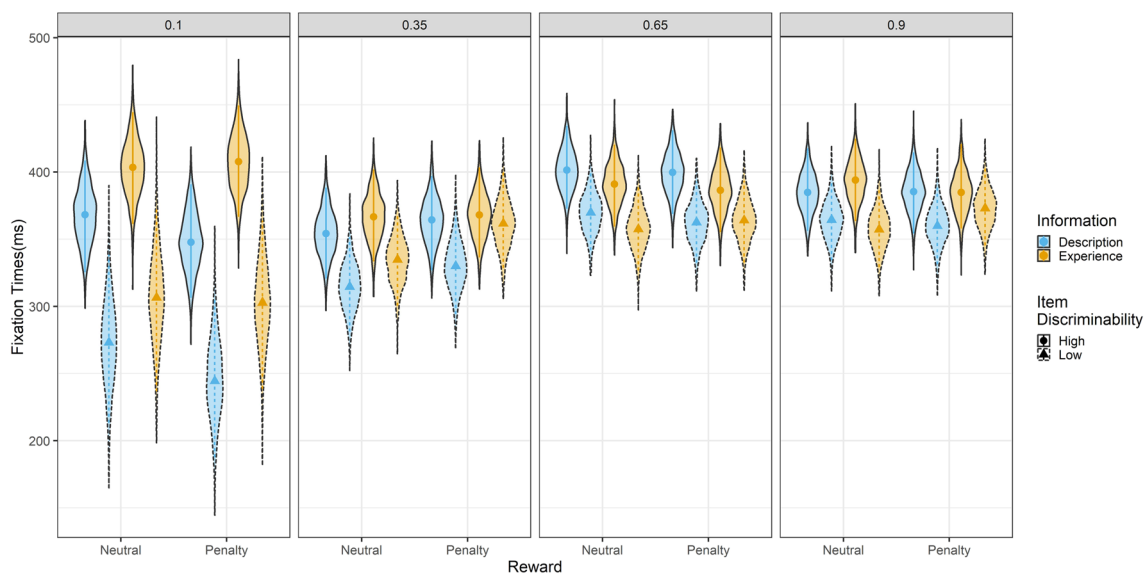


**Fig. 10** The posterior distribution of the group-level target fixation duration

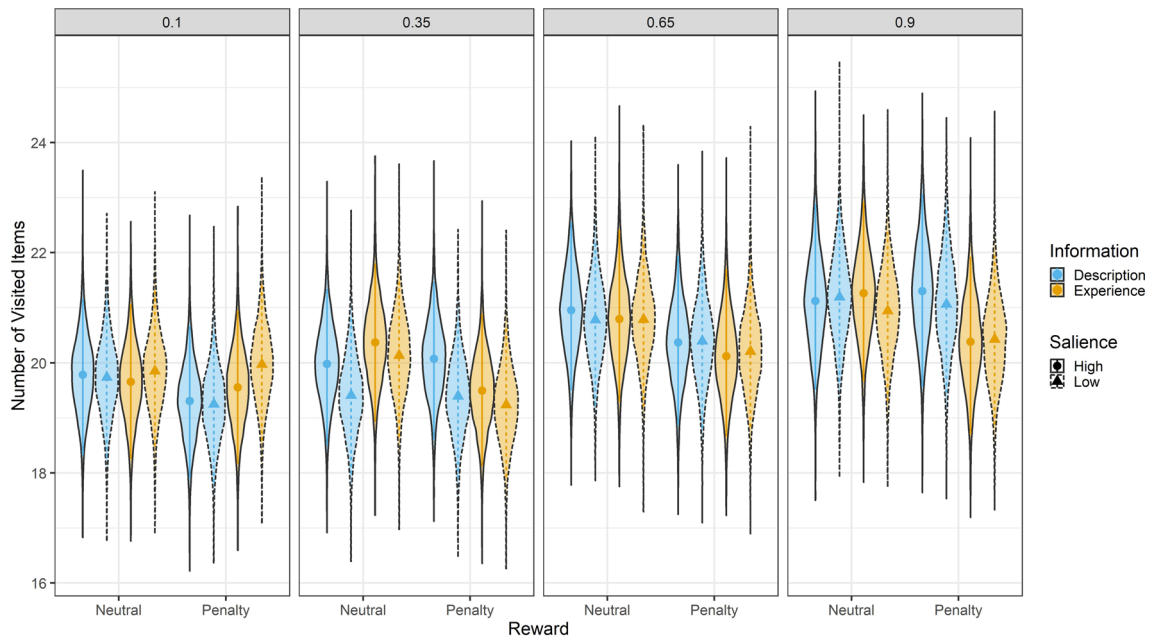
(Bürkner 2017, 2018). Following the analysis procedure from the BayesFactor package (Morey & Rouder, 2018), we computed models that included all possible combinations of manipulations and compared them to the baseline model that only included subject variance. The best model included only main effects of prevalence and reward levels ( $BF = 3.19 \times 10^{21}$ ). There was very strong evidence ( $BF = 49.46$ ) relative to the next best model, which included an additional main effect of salience and the third best model ( $BF = 66.09$ ), which included an additional main effect of information level. For re-fixations items, the best model included all main

effects—information, reward, prevalence, salience—with the three-way interaction of information, reward, and prevalence and the two-way interaction between each two of them ( $BF > 3.59 \times 10^{305}$ ). There was strong evidence of it against the next best model ( $BF = 14.31$ ), which included the additional interaction between salience and information, and stronger evidence ( $BF = 22.28$ ) relative to the third best model, which included the interaction between salience and reward.

Figures 12 and 13 reveal the posterior means and HDIs from the full model including all factors and interactions for the number of visited and revisited items. Our



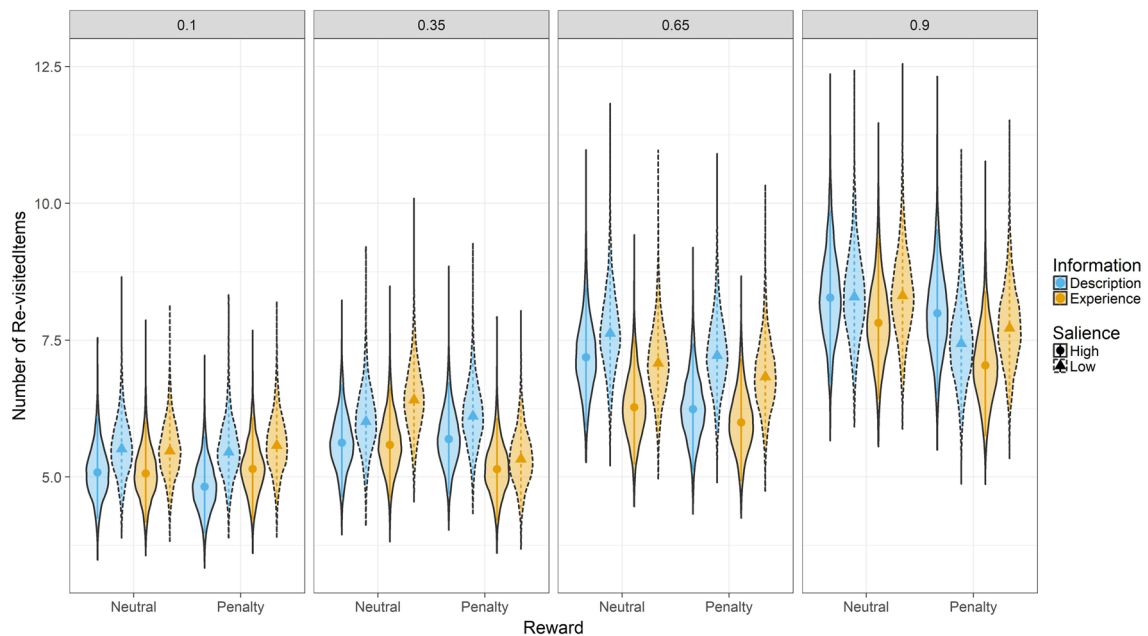
**Fig. 11** The posterior distribution of the group-level distractor fixation duration



**Fig. 12** The posterior distribution of the group-level number of fixated items in *target-absent* trials

analysis showed the basic effect of prevalence on both number of visitations. However, for the number of visited items, there was no other effects other than the prevalence except for the reward manipulation. Our findings in the number of visited items may be due to that observers already put much effort into examining each item—note that there were 25 items in each search field and about 20 of them were visited. The results of revisited items indicated a more interesting pattern: the three- and

two-way interactions in our manipulations of information, reward, and prevalence. Therefore, we find cross interaction between the prevalence effect and information in the re-visited items: Fig. 13 shows that “description” information produced more number re-visited items when the target prevalence is high. More importantly, there were other factors that influence prevalence effect on revisited items for more detailed future exploration.



**Fig. 13** The posterior distribution of the group-level number of re-fixations *target-absent* trials

## Discussion

We systematically examined whether receiving explicit information about prevalence or learning prevalence from experience had an influence on observers' performance. Consistent with most previous research on prevalence effects in visual search (e.g., Wolfe and Van Wert (2010) and Ishibashi et al. (2012)), we found that observers become more biased towards answering target-present as target prevalence increased from low to high. Furthermore, we found a pattern suggesting a “description-experience gap,” which is frequently studied in decision-making literature, but has not yet been examined in the context of visual search: Observers in our study were more biased toward target-present responses with low target prevalence when they learned prevalence from experience compared to when they were informed of the prevalence; whereas in the high prevalence condition, observers were more biased toward target-present responses when they were informed of the prevalence compared to when they learned the prevalence from experience.

Our core analysis was based on a hierarchical Bayesian signal detection model, which we used to estimate biases and subjective prevalence levels, but we also examined the variation using a number of measures commonly used in the prevalence visual search literature. The bias parameter in the signal detection model clearly showed the cross-over interaction between the way prevalence was communicated and the block prevalence indicative of the description-experience gap. Target-absent response time patterns, which reflect how quickly an observer is willing to stop searching for a target (Wolfe et al., 2013), showed a similar pattern to the signal detection analysis. As prevalence increased, observers were slower to respond target-absent when targets were more prevalent and there was a cross-over interaction such that when prevalence was low observers were slower in the experience condition than the description condition, but when prevalence was high observers were slower in the description condition.

The eye-tracking also indicated increased bias toward target-present response with increased prevalence. Observers spent less time fixating targets with the increased prevalence conditions and more time fixated on distractors, consistent with Peltier and Becker (2016). There was not as clear evidence for the cross-over interaction associated with a “description-experience gap”, however fixation times on targets were slightly longer in the experience condition. Both the total number of fixated items and the number of re-fixated items increased with increased prevalence, following the signal detection and RT patterns and similar to the findings of Godwin et al. (2015). There was no

evidence of a description-experience gap in the number of fixated items, but the number of re-fixations did indicate a cross-over interaction.

Within the decision-making literature, three categories of explanation have been posited for the cause of the description-experience gap: While making choices based on experience, decision-makers may explore only a small sample of trials from the whole distribution, focus more on the recent observed outcomes or subjectively weight the probability in a different way (Wulff et al., 2018; Hertwig & Erev, 2009). These mechanisms can lead to the decision-experience gap because decision-making behavior differs from classic description-based paradigm in which preference is measured for a choice once and in isolation (e.g., Tversky and Kahneman (1992)). Results from previous visual search studies indicate that both small-samples and recency can influence prevalence effects in experience-based prevalence search. For example, Horowitz (2017) reviewed previous research and noted that knowledge of target prevalence develops over a window of approximately 20-50 trials. Thus, observers may rely on the “local prevalence” to expect whether the target shows up subsequently or not (e.g., Ishibashi et al. (2012) and Wolfe and Van Wert (2010)). Horowitz (2017) also proposed that the discrepancy in findings about prevalence effect between medical image perception (e.g., Reed et al. 2011, 2014) and other psychological studies (e.g., Lau and Huang (2010)) was that the small samples in medical image perception studies were not sufficient for building explicit expectations. Our current results on the description-experience gap add additional explanations on the prevalence effect, showing that observers' subjective interpretations on target-probability can play a role in addition to the controlled effects of “small samples” and “recency”. Therefore, future research on prevalence effect should investigate probability learning (e.g., Estes (1976)) in further as well as other effects such as “small sample” and “recency” that might lead to the prevalence effect.

In addition to the “description-experience” gap, one finding that diverged from previous results was the degree to which participants underestimated the probability of a target across a wide range of prevalence levels and conditions. In the experiment settings, we adopted the manipulation of payoff rewards that associated with different responses. Our results indicated that observers behaved differently when there was an extreme penalty with “missed target”, e.g., when the target prevalence was high, observers were much more likely to respond “target-present” based on the criteria in signal detection analysis. However, in contrast with Navalpakkam et al. (2009), observers were not able to obtain the optimal criteria except the “neutral” reward

in 0.1 condition (more details see Appendix A). One possible reason was that our participants were generally inexperienced (mostly undergraduate students) relative to expert radiologists. Naïve observers who are unfamiliar with the search task may be more likely to be influenced by the experienced target-present event. Expert radiologists, on the other hand, have a thorough understanding on the target they are searching for and could more influenced by the given explicit information. A potential future direction related to this finding is to explore the interaction between observers' search experience with information discrepancy given by the task. Our task also measures the difficulty of stimuli for subjects to identify, which can also speak to the proposed decision process model by Peltier and Becker (2016).

## Conclusions

Our current study examined the influence of reward schemes and information communication on observers' search performance when observers were instructed to search for a target in different target prevalence contexts. Our results indicated that searches based on explicit target prevalence knowledge differed from searches based on accumulated information gained through experience. When observers search in high prevalence conditions based on explicit prevalence instructions, they are more biased toward target-present response than when they search based on experienced prevalence while it is the reverse when the target prevalence is low. While there is a significant body of research on prevalence effects in visual search, we believe this is the first paper to connect these effects to the description-experience gap in the decision-making literature.

**Open practice statement** This study was not preregistered. The data and analysis scripts are available upon request.

## Appendix A: An optimal criteria in visual search

Signal detection theory (SDT; Green and Swets (1966)) assumes that the internal response to a target/distractor is sampled from an equal variance Gaussian distribution (Wickens, 2002). The perceptual discriminability ( $d'$ ) between the two distributions determines how well observers can discriminate targets from distractors—the larger  $d'$  represents that observers can identify the targets better. By setting up a perceptual criterion ( $\lambda$ ), responses

can be categorized into four different types (Table 2): correctly identify a target (*hit*), falsely identify a distractor as the target (*false alarm*), correctly reject a distract (*correct rejection*), and fail to report a target (*miss*).

The dual-threshold model (Wolfe & Van Wert, 2010) assumes that the criteria  $c$  shifts, where  $c$  is the distance between  $\lambda$  and the center of the two distributions,  $d'/2$  ( $\lambda_{center}$ ):

$$c = -\frac{z(H) + z(FA)}{2} = \lambda - \frac{1}{2}d'$$

Therefore,  $c$  represents the relative bias of responding “yes” (target-present) or “no” (target-absent), when  $c = 0$ , the bias is equal. To optimize the accuracy, the ratio of responding “yes” or “no” should depend on the likelihood of target present  $s$ :

$$c_{optimal} = -\frac{\text{logit}(s)}{d'}$$

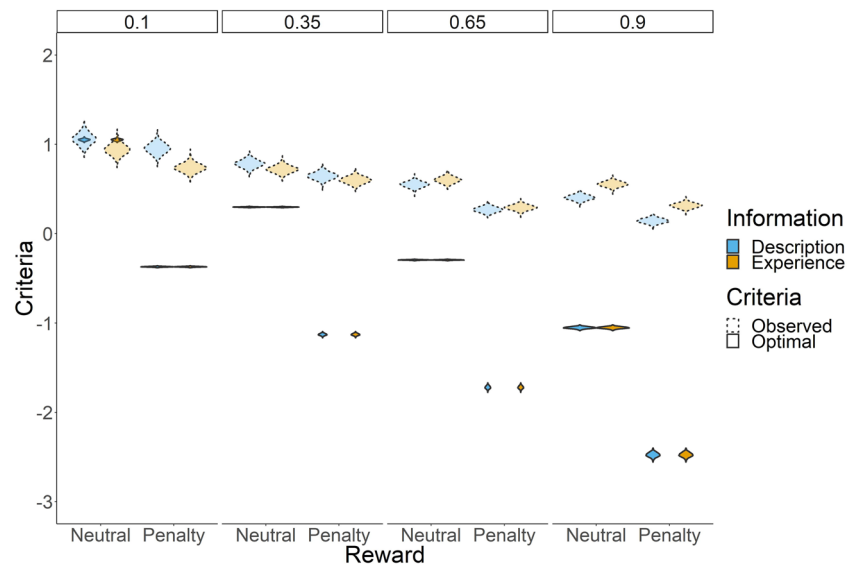
In a visual search study, when the target prevalence is 50%, the optimal strategy to maximize the accuracy is to posit the criteria  $c$  at the center of the two distributions where observers are expected to respond “yes” and “no” equally. When the target prevalence drops down to a low prevalence condition,  $c_{optimal}$  is biased towards “no”, thus producing more target-absent responses as described in the dual threshold model (Wolfe & Van Wert, 2010). To persuade more target-present responses, our current study employed the reward schemes from the previous study by Navalpakkam et al. (2009), thus instead of trying to maximize the accuracy, the observers were instructed to maximize the payoff, which was given by

$$c_{optimal} = \frac{1}{d'} \left[ \log \frac{V(CR) - V(FA)}{V(H) - V(M)} - \text{logit}(s) \right]$$

As Table 1 indicates, in the *neutral* condition,  $[V(CR) - V(FA)]/[V(H) - V(M)] = 1$ . Thus  $c_{optimal}$  is obtained by  $-\text{logit}(s)/d'$ . Observers are expected to shift their criteria  $c$  accordingly with the target prevalence  $s$  changes as observed in many previous studies (e.g., Ishibashi et al. (2012)). In the *penalty* condition,  $[V(CR) - V(FA)]/[V(H) - V(M)] = 0.051$ , the influence of target prevalence  $s$  on  $c_{optimal}$  is attenuated by the reward scheme. Thus, we hypothesized that observers were biased towards responding target-absent to ameliorate the prevalence effect as observed in Navalpakkam et al. (2009).

**Table 2** Responses in signal detection theory

Responses	Target	
	Present	Absent
Present	Hit (H)	False alarm (FA)
Absent	Miss (M)	Correct rejection (CR)



**Fig. 14** Violin plot depiction of observed criteria and optimal criteria, sampled from the posterior distributions in the Bayesian SDT

As indicated in Fig. 14, observers shifted their criteria as the penalty on miss errors, which was consistent with the Navalpakkam et al. (2009). However, opposite from the arguments by Navalpakkam et al. (2009), the penalty condition created an even harder scenario for observers to catch the optimal criteria. Instead, only part of the posterior distribution in the neutral reward condition when the prevalence was low captured the optimal criteria.

## References

- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395–411.
- Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, 83(1), 37–64.
- 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.
- Fox, C. R., & Hadar, L. (2006). Decisions from experience"= sampling error+ prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, 1(2), 159–161.
- Godwin, H. J., Menneer, T., Cave, K. R., Thaibsyah, M., & Donnelly, N. (2015). The effects of increasing target prevalence on information processing during visual search. *Psychonomic Bulletin & Review*, 22(2), 469–475.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Gronau, Q. F., & Singmann, H. (2018). bridgesampling: Bridge sampling for marginal likelihoods and Bayes factors [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=bridgesampling> (R package version 0.6-0).
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8), 534–539.
- Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, 13(12), 517–523.
- Horowitz, T. S. (2017). Prevalence in visual search: from the clinic to the lab and back again. *Japanese Psychological Research*, 59(2), 65–108.
- Ishibashi, K., & Kita, S. (2014). Probability cueing influences miss rate and decision criterion in visual searches. *i-Perception*, 5(3), 170–175.
- Ishibashi, K., Kita, S., & Wolfe, J. M. (2012). The effects of local prevalence and explicit expectations on search termination times. *Attention, Perception, & Psychophysics*, 74(1), 115–123.
- Jeffreys, H. (1961). *Theory of probability*, (3rd ed.). Oxford: Oxford University Press.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decisions under risk. *Econometrica*, 47, 263–292.
- Kundel, H. L. (1982). Disease prevalence and radiological decision making. *Investigative Radiology*, 17(1), 107–109.
- Kundel, H. L. (2000). Disease prevalence and the index of detectability: a survey of studies of lung cancer detection by chest radiography. In *Proceedings of SPIE: Image perception and performance*, (Vol. 3981, pp. 135–145).
- Lau, J. S.-H., & Huang, L. (2010). The prevalence effect is determined by past experience, not future prospects. *Vision Research*, 50(15), 1469–1474.
- Lee, M. D. (2018). Bayesian methods in cognitive modeling. In Wagenmakers, J. E.-J. (Ed.) *Stevens' handbook of experimental psychology and cognitive neuroscience*. (4th ed., Vol. 5, pp. 37–84). New York: Wiley.
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.
- Littlefair, S., Brennan, P., Reed, W., & Mello-Thoms, C. (2017). Does expectation of abnormality affect the search pattern of radiologists when looking for pulmonary nodules? *Journal of Digital Imaging*, 30(1), 55–62.
- Littlefair, S., Mello-Thoms, C., Reed, W., Pietryzk, M., Lewis, S., McEntee, M., et al. (2016). Increasing prevalence expectation in thoracic radiology leads to overcall. *Academic Radiology*, 23(3), 284–289.



- McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. London: Chapman and Hall/CRC.
- Meredith, M., & Kruschke, J. (2018). HDInterval: Highest (posterior) density intervals [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=HDInterval> (R package version 0.2.0).
- Morey, R. D., & Rouder, J. N. (2018). Bayesfactor: Computation of Bayes factors for common designs [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=BayesFactor> (R package version 0.9.12-4.2).
- Navalpakkam, V., Koch, C., & Perona, P. (2009). Homo economicus in visual search. *Journal of Vision*, 9(1), <https://doi.org/doi:10.1167/9.1.31>.
- Nocum, D. J., Brennan, P. C., Huang, R. T., & Reed, W. M. (2013). The effect of abnormality-prevalence expectation on naïve observer performance and visual search. *Radiography*, 19(3), 196–199.
- Peirce, J. W. (2007). PsychoPy – psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8–13.
- Peltier, C., & Becker, M. W. (2016). Decision processes in visual search as a function of target prevalence. *Journal of Experimental Psychology: Human Perception and Performance*, 42(9), 1466–1476.
- Peltier, C., & Becker, M. W. (2017a). Individual differences predict low prevalence visual search performance. *Cognitive Research: Principles and Implications*, 2(1), <https://doi.org/10.1186/s41235-016-0042-3>.
- Peltier, C., & Becker, M. W. (2017b). Target-present guessing as a function of target prevalence and accumulated information in visual search. *Attention, Perception, & Psychophysics*, 79(4), 1064–1069.
- Reed, W. M., Chow, S. L. C., Chew, L. E., & Brennan, P. C. (2014). Can prevalence expectations drive radiologists' behavior? *Academic Radiology*, 21(4), 450–456.
- Reed, W. M., Ryan, J. T., McEntee, M. F., Evanoff, M. G., & Brennan, P. C. (2011). The effect of abnormality-prevalence expectation on expert observer performance and visual search. *Radiology*, 258(3), 938–943.
- 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), <https://doi.org/10.1167/8.15.15>.
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12(4), 573–604.
- Schwark, J. D., MacDonald, J., Sandry, J., & Dolgov, I. (2013). Prevalence-based decisions undermine visual search. *Visual Cognition*, 21(5), 541–568.
- Schwark, J. D., Sandry, J., MacDonald, J., & Dolgov, I. (2012). False feedback increases detection of low-prevalence targets in visual search. *Attention, Perception, & Psychophysics*, 74(8), 1583–1589.
- Smith, P. L., & Little, D. R. (2018). Small is beautiful: in defense of the small-n design. *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
- Stan Development Team (2018). RStan: the R interface to Stan. Retrieved from <http://mc-stan.org/> (R package version 2.17.3).
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.
- 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.
- Wickens, T. D. (2002). *Elementary signal detection theory*. USA: Oxford University Press.
- 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), <https://doi.org/10.1167/13.3.33>.
- Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: 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), 623–638.
- 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.
- Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, 144(2), 140–176.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.