An Optimal Foraging Model of Human Visual Search

Matthew S. Cain (matthew.s.cain@duke.edu)
Kait Clark (kait.clark@duke.edu)
Stephen R. Mitroff (mitroff@duke.edu)
Duke University Center for Cognitive Neuroscience & Department of Psychology & Neuroscience,
B203 LSRC, Box 90999, Durham, NC 27708 USA

Ed Vul (evul@ucsd.edu)
Department of Psychology, University of California, San Diego
9500 Gilman Dr., La Jolla, CA 92093

Abstract

Unlike laboratory experiments, real-world visual search can contain multiple targets. Searching for an unknown number of targets creates a unique set of challenges for the observer, and often produces serious errors. We propose a Bayesian optimal foraging model to predict and describe behavior in such search scenarios, and investigate whether people adapt their search strategies based on complex statistics of target distributions. Separate groups searched arrays drawn from three target distributions with the same average number of targets per display, but different target-clustering properties. As predicted, participants searched longer when they expected more targets to remain and adjusted their expectations as searches unfolded, indicating that searchers are sensitive to the target distribution, consistent with both an optimal foraging framework and an ideal Bayesian observer. However, compared to the ideal observers, searchers systematically under-adjusted to the target distribution, suggesting that training could improve multiple-target search in radiology and other crucial applications.

Keywords: Visual search; Optimal foraging theory; Bayesian modeling.

Introduction

When should a radiologist stop searching for abnormalities on your X-Rays? How long should an airport baggage screening officer search through your bag? Searching for important objects in clutter is a ubiquitous real-world task, which has been most systematically studied in vision (see Nakayama & Martini, 2010, for a review). The bulk of this literature focuses on search for a single target (e.g., looking for your keys); however, some of the most important real-world visual search tasks (e.g., those conducted by radiologists, baggage screeners, and military personnel) are multiple-target searches—searches where there might be many targets in a given display (e.g., an X-ray can contain an unknown and unbounded number of potential abnormalities). This presents an interesting problem to the searcher: when to stop searching?

It is sometimes possible to search exhaustively by inspecting every potential target; however, in most situations this is prohibitively costly and inefficient. When the number of possible targets is unknown, searching efficiently requires tailoring a stopping decision to the expected target distribution. Evidence suggests that people do adapt their search behavior to environmental statistics. Multiple-target search errors reflect, in part, the probability of a target being present, with less frequent targets being missed more often (e.g., Fleck, Samei, & Mitroff, 2010). Similarly, prevalence in single-target search influences decision criteria, resulting in more false alarms at high target prevalence and more misses at low target prevalence (Godwin, Menneer, Cave, & Donnelly, 2010; Wolfe & Van Wert, 2010). Moreover, these prevalence effects may be driven by experiences prior to the immediate task (Lau & Huang, 2010). Although these data suggest that human search behavior adapts to environmental statistics, there has been no test of the sophistication of these adaptations, nor is there a theory to explain how people decide to stop searching when the number of targets is unknown. Here we propose two related models of human stopping behavior in multi-target search and test whether people are as sensitive to complex target distributions as these models.

Figure 1: A sample search array. All displays contained 40 items, 0–12 of which were targets (perfectly aligned T shapes, 6 present here).

Optimal Searcher Models

We expect common, effortful, cognitive tasks to efficiently exploit environmental statistics to improve performance. While there has been little work in this area for human search, there is an extensive animal literature on optimal foraging theory that has formalized how animals may make use of the statistical properties of their environments (see Stephens & Krebs, 1986, for a review). Optimal foraging is typically used to explain the behavior of animals consuming food and to answer questions such as "when should a bird eating berries off
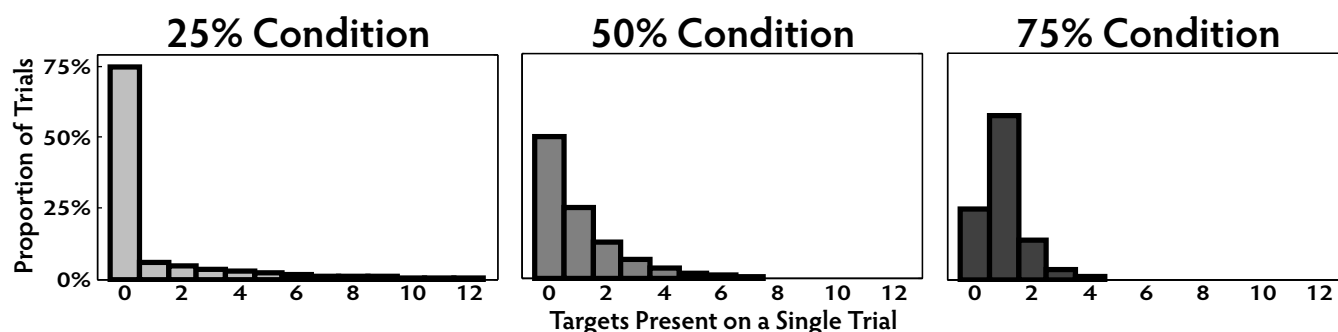
Figure 2: Histograms of the number of targets present in each display. Each participant saw a distribution of targets drawn from a truncated geometric distribution with an expected value of 1 target per trial, but with groups differing on the proportion of trials that had targets.

a bush move on to another bush?" The key insight of optimal foraging theory is that an organism should strive to maximize its rate of energy intake—rather than, say, ensuring that it has consumed all available food—and should leave one patch and move to the next when the instantaneous rate of energy intake for the current patch falls below the expected rate for the environment as a whole. In short, a forager should aim to spend time in above-average patches and leave them before they drop below average. While this technique has been used almost exclusively to analyze animal behavior, a recent study of human vigilance employing such a model found that humans were sensitive to patchy, negative binomial distributions of stimuli (Hutchinson, Wilke, & Todd, 2007). Moreover, optimal foraging theory can be adapted into a Bayesian ideal observer model, where the instantaneous rate of return is estimated based on a continuously updated posterior over the value of the search area.

We investigate whether people adjust their search strategy in a manner consistent with these optimal foraging/search models. Three groups of participants searched for visual targets under different distributions of the number of targets present per display. Each group saw, on average, one target per display, but targets were 'clustered' differently across conditions. In one condition, only 25% of trials contained targets, but those trials tended to contain many targets, whereas in the other extreme condition, 75% of trials contained targets, but those trials tended to contain only one target each. Do human search strategies adapt to these manipulations of higher-order target distribution statistics?

Methods

Experiment

Participants Forty-five members of the Duke University community (28 female; aged 18 to 48 years, median = 23) were asked to find 'T's (characterized by a perfectly bisecting line) among 'L's (imperfectly bisecting line; see Figure 1). Participants were awarded 15 points for each target found and the experiment ended when they reached 2000 points.[1]

[1]These selected values were based on pilot testing to ensure that the experimental session would not exceed 1 hour.

There were no penalties for misses or false alarms.

Critically, participants were randomly assigned to one of three between-participants conditions that manipulated the target distributions while holding constant the expected number of targets per trial at a value of one (Figure 2). The number of targets present in each trial was sampled from a geometric distribution with the rate parameter adjusted per condition to yield the same average number of targets per trial. In the 25% condition, only one-quarter of the trials had at least one target, but those trials were likely to contain many targets. In the 50% condition, half of the trials had at least one target and of those trials, half had one target and half had more than one. In the 75% condition, three-quarters of the trials had at least one target, but those trials were unlikely to contain more than one target. These target distributions provided complex, but informative, target prevalence statistics which could be exploited to adjust the rate of experiment attrition.

Procedure Prior to the main experiment, there was a practice block with a 120-point goal to familiarize participants with the experiment and the target distribution (which was matched to their condition). Each trial began with a cross appearing for 0.5 s at the center of the screen. The cross was replaced with a search array of 40 grey items on a cloudy grey background (Figure 1), targets and distracters were randomly positioned within the search array. Items randomly varied between 27–65% black. Participants clicked on each T they found (with the location of the click marked with a small blue circle) then clicked a button labeled "Done", ending the trial. Feedback after each trial revealed all the targets that were present in order to provide all participants the same information about the target distribution, regardless of their performance.

Predictions

How would people perform if they adjusted their search strategy based on target distribution statistics? We can formalize our predictions via a Bayesian ideal observer that starts every trial with a prior over the number of targets ($P(T)$) matched to the distribution of targets across displays, and computes a posterior ($P(T|F,S)$), after having found F targets having
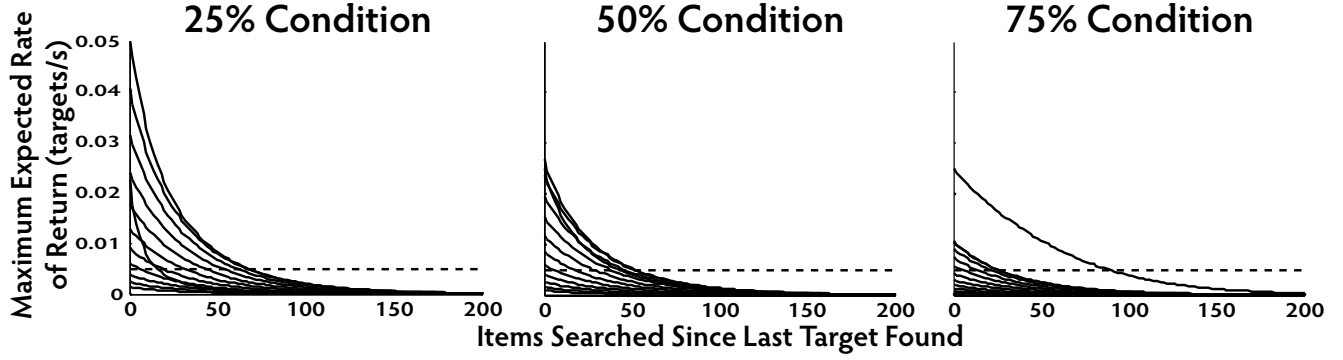
Figure 3: Expected rates of return for the different conditions as a function of the number of targets found (different lines, from 0 to 12), and the number of items searched since the last target was found. The dashed horizontal line in each plot corresponds to the maximum expected rate of return of switching to a new trial. The ideal observer should continue searching on the current trial so long as the instantaneous expected rate of return exceeds the maximum expected rate of return of switching.

searched S items: $P(T|F,S) \propto P(F|T,S)P(T)$.

For $P(F|T,S)$ we used a sampling distribution without replacement of targets: When a target is found, it cannot be marked again, but when a distracter is found, it continues to be sampled during the rest of the search (Horowitz & Wolfe, 2001). Thus, when searching for 5 targets among 40 items, the probability of finding a target is 5/40 on each draw until the first target is found, then the probability of finding the second target on subsequent draws becomes 4/39, etc. The pseudo-hypergeometric distribution that results from such half-replacement has no analytical form, but it can be numerically calculated with high precision for the range of F, S, and T values we use here.

The posterior over the number of targets yields a predictive distribution: the probability of finding K more targets when searching R more items by marginalizing over all possible total numbers of targets.

$$P(K|R,F,S) = \sum_T (P(K|T,R)P(T|F,S))$$

where $P(K|T,R)$ corresponds to the same distribution as used for $P(F|T,S)$. This predictive distribution defines an expected return rate arising from searching R more items:

$$\frac{E[K|R,F,S]}{R} = \frac{\sum_K (K * P(K|R,F,S))}{R}$$

Since the probability of finding another target after sampling one more item is less than 1, $E[K|R,F,S]$ increases more slowly than R; thus the peak value is always at $R = 1$, meaning that the anticipated instantaneous rate of return corresponds to the maximum expected rate of return. And just as in optimal foraging theory, the ideal observer will switch trials when the instantaneous expected rate of return for the current trial falls below the maximum expected rate of return of switching to a new trial.

The expected rate of return for switching to a new trial can be calculated similarly, based only on the prior over the target distribution:

$$\frac{E[K|R]}{R+\tau} = \frac{\sum_{K,T}(K \times P(K|T,R)P(T))}{R+\tau}$$

where τ reflects the forgone search opportunity during the 3s period between trials. Unlike the expected instantaneous rate of return, $E[K|R]/(R+\tau)$ has a peak value at an $R > 1$. Figure 3 shows the expected instantaneous rates of return for each number of targets found for each condition, as a function of time spent searching after the last identified target was found. When the expected rates of return fall below the maximum expected rate of return of switching to a new trial, the ideal observer should end the trial.

The ideal Bayesian observer captures the intuitive prediction (Figure 4B): in the 25% condition, before the first target is found, the trial is likely to contain few targets, so it does not make sense to search it for a long time, but once a single target is found, many more are likely, so it would be best to continue searching for a longer period of time. In contrast, in the 75% condition, a target is likely to be present in the display, but once it is found, a second one is not likely. This crossover is a characteristic of optimally using these target distributions during search.

Results

In order to assess how valuable searchers considered a given display, we used the amount of time spent searching after finding a target as a proxy for display quality, with participants assumed to be willing to search longer in displays that they determined were likely to contain more targets. Specifically, we measured the time between finding the last target actually found in a display (even if not all targets were found) and clicking Done to end the search. Figure 4A plots this difference measure for each condition against the number of targets found. The main pattern is clear: the more targets that were likely to be in a display, the longer participants continued searching before terminating their search. This was confirmed with a 3×4 mixed model ANOVA with condition (25%, 50%, or 75%) and number of targets found (0,
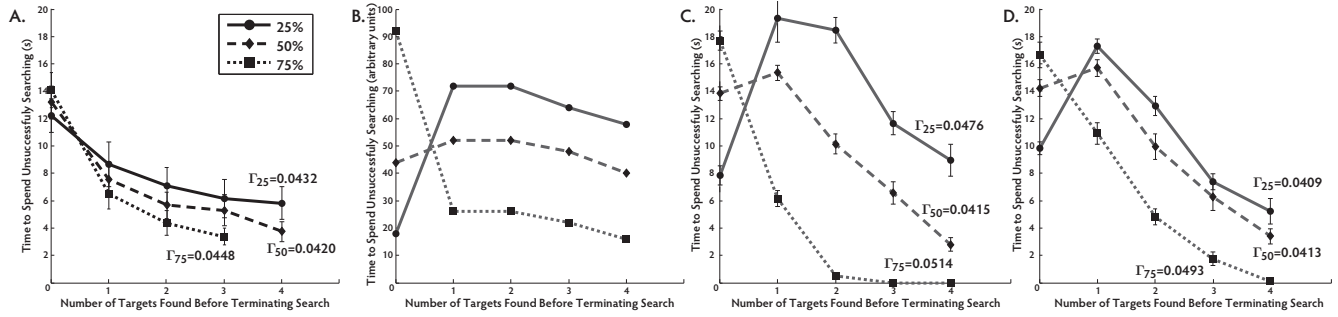
Figure 4: (A) The time spent searching after finding the last target found in a particular display, plotted as a function of the number of targets found (not necessarily all the targets present) for each condition. The value for 4 targets is not plotted for the 75% condition, as fewer than half the participants in this condition found 4 targets in a single display. (B) Predictions of a Bayesian ideal observer. When 0 targets have been found, observers in the 75% condition have cause for optimism, and are expected to search for a long time before stopping; while those in the 25% condition should be pessimistic about the prospects of the display containing any targets at all. However, once a target has been found, these expectations reverse: the observers in the 75% condition should not expect to find any more targets, while observers in the 25% condition should expect to find many more. This should result in participants in the 25% condition searching longer for the 2nd, 3rd, 4th, target, etc. (C) Potential Value theorem model results with an optimal Bayesian value threshold. (D) Model results assuming only partial (50%) learning of the target distributions.

1, 2, or 3)[2] as factors. There were significant main effects of condition ($F(2, 126) = 4.86, p = 0.013$) and of number of targets found ($F(3, 126) = 161.48, p < 0.001$) and a significant interaction between the variables ($F(6, 126) = 5.08, p < 0.001$).

This interaction confirms the qualitative predictions of the Bayesian ideal observer: finding the first target made people search longer in the 25% condition, but made them stop earlier in the 75% condition. However, there is a substantial deviation between the magnitude of the predicted and observed effects. It seems that despite adapting to the target distribution statistics, as predicted, participants did not optimally determine their stopping times, but tended to search overly long when no targets had been found and in the 25% condition and not long enough when 1 target had been found in the 25% and 50% conditions.

Potential Value Model

The marginal value theorem (Charnov, 1976), states that an optimal forager should abandon search at the current location when the rate of return of the current location reaches the rate for the environment. The potential value theorem (PVT; McNamara, 1982) extends the marginal value theorem via a stopping rule that takes into account information gained by finding a target. Like the marginal value theorem, the PVT assumes that searchers are aware of the overall distribution of targets in their environments, their overall target encounter rate, and their target encounter rate for the current trial. Here we employ a Bayesian potential value rule (Olsson & Brown, 2006) that builds upon the Bayesian ideal observer model described previously.

Overall Rate of Target Finding   Searchers are presumed to maximize Γ, their rate of finding targets across the whole

experiment (or in our case, minimizing the total time to accumulate 2000 points and finish the experiment):

$$\Gamma = \frac{\bar{n}}{\bar{t} + \tau}$$

where $\bar{n}$ is the average number of targets found per trial, $\bar{t}$ is the average time spent searching per trial and τ is the constant inter-trial interval. A higher Γ indicates that the searcher spent less time searching fruitless displays and more time searching trials with many targets. Guiding behavior to maximize Γ requires estimating two values: the current trial's quality (updated continuously) and the quality threshold below which the searcher should switch trials (a constant value for a given searcher).

We calculated Γ—the overall rate at which targets were found—for each searcher. A 1-way ANOVA with condition as a factor revealed that Γ did not differ between distribution conditions ($\Gamma_{25} = 0.045 \pm 0.008$ targets/s, $\Gamma_{50} = 0.042 \pm 0.006$ targets/s, $\Gamma_{75} = 0.044 \pm 0.010$ targets/s; $F(2, 42) = 0.44, p = 0.647$). Despite searching under different target distributions, each group settled on the same average target acquisition rate. Such a result would not be predicted by simple foraging models such as fixed search time models, and strongly suggests that searchers were adapting their behavior to the statistics of their search environment in order to achieve a high rate of target detection.

Estimated Trial Quality   Under the PVT, trial quality— how much a trial is worth to the searcher—is estimated in targets per second, and denoted as Π:

$$\Pi = \frac{E(n)}{E(t)}$$

where $E(n)$ is the expected number of targets remaining in the display, and $E(t)$ is the how long it is expected to take to find them. We computed the estimated trial quality, Π,

---

[2]Fewer than half of the participants in the 75% condition found more than 3 targets on any trial.

at each 500ms intervals of each search trial. $E(n)$ and $E(t)$ were computed based on the number of targets found and the amount of time already spent searching using the ideal observer model described above. Figure 5 shows Π as it unfolds over a representative trial from each condition. The estimated trial quality decays exponentially as search continues, but changes discontinuously when a target is found. These changes represent the information gained by finding a target. For example, in the 25% condition, before the first target is found, knowledge of the distribution gives a 0.75 likelihood of the display containing zero targets, and a 0.25 likelihood of the display containing one or more targets. Once a target is found, however, the chance of the display containing at least one additional target jumps to 0.75. This additional information accounts for the increase in estimated quality of the 25 % condition trial in Figure 5 seen at about 2s. In contrast, for the 75% condition, finding a target actually *decreases* the estimated trial value: half of the displays in the 75% condition have exactly one target, so finding one target indicates that finding another target in that display is unlikely. This can be seen Figure 5 in the large drop in estimated quality for the 75% condition trial at about 3s.

For each participant we calculated the estimated trial quality at the time of search termination. These values were submitted to the same 3×4 mixed model ANOVA as the quitting times, above. There was not a significant main effect of condition ($F(2, 126) = 2.37$, $p = 0.106$), but there was a significant main effect of number of targets found ($F(3, 126) = 114.52$, $p < 0.001$) and a significant interaction between the factors ($F(6, 126) = 172.23$, $p < 0.001$). While there was no overall difference across conditions, posthoc one-way ANOVAs revealed that condition differences were significant at all four levels of number of targets found (0: $F(2, 42) = 196.88$, $p < 0.001$; 1: $F(2, 42) = 7.30$, $p = 0.002$; 2: $F(2, 42) = 12.82$, $p < 0.001$; 3: $F(2, 42) = 11.56$, $p < 0.001$; Šidák-corrected threshold=0.010). This suggests that participants are adjusting to the target distribution statistics, in terms of their rate of target discovery, Γ, and their raw stopping times, by finding very different quitting thresholds for each condition. Interestingly, while the PVT implements a set quitting threshold across all trials, the main effect of number of targets found suggests that participants had more variable thresholds.

Optimality of Search Termination

Two components affect when searches are terminated, estimated trial quality and the trial value quitting threshold. Under the potential value theorem each searcher has a fixed quitting threshold across all trials. This threshold determines how many targets are found and how much time is spent searching. An extreme threshold of very near zero would indicate exhaustive search with all targets found after a long time searching, while a very high threshold would result in search being terminated very quickly, without finding any targets. The optimal threshold, Π^*, should be between these extremes and should be sensitive to the characteristics of the environment.
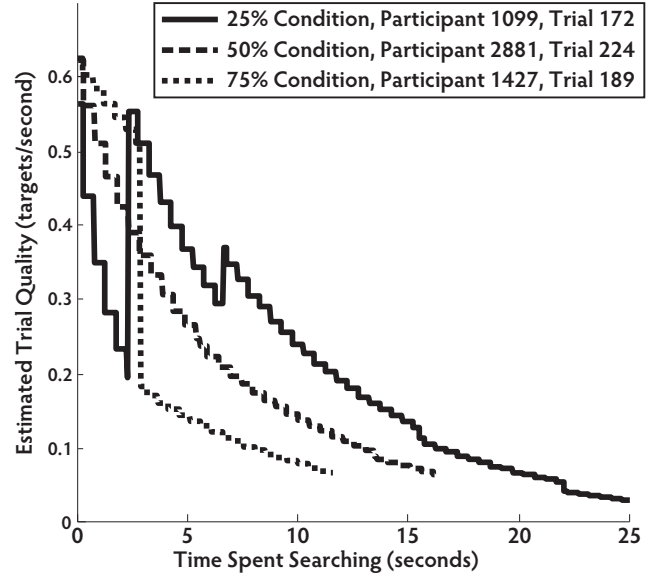

Figure 5: Estimated trial quality, Π, over time for three representative trials. Note that finding a target increased estimated trial quality at the beginning of the trial in the 25% and 50% conditions but decreased it toward the end of those trials and immediately in the 75% condition.

In order to find Π^* we examined the actual search timecourses for each trial and tallied the number of targets that would have been found and amount of time that would have been spent searching over a large range of quitting thresholds. To account for trials in which the actual search terminated before the candidate quitting threshold had been reached, Π was calculated out to 90s of simulated searching, with random target finding events generated based on each participant's overall search speed and the remaining target density in the display. The Γ values for each candidate quitting threshold were compared and the threshold that produced the largest Γ for each searcher was deemed to be their Π^*. Group averages of Π^* and their resulting Γs are shown in Figure 4C.

The clearest result is that searchers in the 75% condition could have done much better by searching longer when no targets had been found and then giving up immediately upon finding a second target ($\Gamma_{observed} = 0.0448 \pm 0.008$; $\Gamma_{\Pi^*} = 0.0514 \pm 0.008$; $t(28) = 2.24, p = 0.033$). Conversely, searchers in the 25% condition could have done better by quitting their search more quickly before finding a target but searching longer after the first target was found ($\Gamma_{observed} = 0.0447 \pm 0.010$; $\Gamma_{\Pi^*} = 0.0476 \pm 0.010$). This deviation between optimal searching and human behavior is seen when compared to both the Bayesian ideal observer (Figure 4B) as well as PVT threshold model (Figures 4B & C). The searchers in the 50% condition performed quite closely to the Bayesian potential value prediction, slightly, but non-significantly outperforming the model ($\Gamma_{observed} = 0.0420 \pm 0.006$; $\Gamma_{\Pi^*} = 0.0415 \pm 0.004$).

Together, these results suggest that searchers in the 50%

condition are accurately using their expectation about trial quality to influence their performance and adjust these expectations to new target distributions. However, searchers do not adjust their strategies to take sufficient advantage of the more extreme target distributions. One possible explanation is that despite practice, participants may have a strong prior that half of trials ought to have a target—an exceedingly common scenario in psychology experiments. Figure 4D illustrates the effects on the model when the extreme distributions are only 50% learned (i.e., that they are assumed to be less extreme), which is a much closer approximation of the actual data. If such a prior accounts for human deviations from optimality, it may be possible to improve applied multiple-target search performance by training with specially-tuned target distributions.

Discussion

The present experiment demonstrates that searchers adapt in sophisticated ways to the complex statistics of their search environment: participants terminated their searches quickly when finding an additional target was unlikely but searched longer when finding an additional target was more likely. This was observed both within participants (with a main effect of number of targets found), and between groups (with an interaction between condition and number of targets found), suggesting that people optimize their search strategies to the environment. While the searchers in the 50% condition performed nearly optimally, searchers in the 25% and 75% groups deviated from optimal predictions by not adjusting sufficiently to the target distributions.

Only one other study to date that has examined sensitivity to patchy target distributions in a human visual cognition task. Hutchinson et al. (2007) presented participants with a simulated fishing task in which fish appeared at a rate dependent on the number remaining in the pond. Participants could switch to a new pond at any time (with a fixed transition duration). They found that participants generally responded appropriately when presented with clustered targets, but dwelled longer than optimal on a given pond. Our fine-grained analysis shows deviations in both directions from optimality depending on the number of targets found and the target distribution: while searchers adjust to the target distribution, they do not adjust as much as is optimal.

To date, no model of multiple-target visual search has been put forward to explain search termination behavior. Here we propose a class of models that predict how search termination behavior is strategically adjusted based on the expectations for the current and subsequent trials. Our experiment demonstrates that although participants do not adjust their strategies as much as would be optimal, they are sensitive to the same factors used by the ideal observers to determine stopping times. We suspect that the same strategic considerations underlying the behavior we observed likely account for target prevalence (Wolfe & Van Wert, 2010) and satisfaction of search (Fleck et al., 2010) effects in visual search. The

generality of this mechanism has broad implications, suggesting that artificially modifying target distribution statistics, such as priming baggage screeners with daily training runs of multiple-target bags (c.f., Wolfe et al., 2007), may be an effective way to enhance performance in critical multiple-target visual searches and could curtail early termination errors which may be responsible for up to half of missed abnormalities in radiology (Berbaum, Franken, Caldwell, & Schartz, 2010).

References

Berbaum, K. S., Franklin, E. A. Jr., Caldwell, R. T., & Schartz, K. M. (2010). Satisfaction of Search in Traditional Radiographic Imaging. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques*. Cambridge University Press.

Charnov, E. (1976). Optimal foraging: The marginal value theorem. *Theoretical Population Biology*, *9*(2), 129–136.

Fleck, M. S., Samei, E., & Mitroff, S. R. (2010). Generalized 'satisfaction of search': Adverse influences on dual-target search accuracy. *J. of Experimental Psychology: Applied*, *16*(1), 60–71.

Godwin, H., Menneer, T., Cave, K., & Donnelly, N. (2010). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition*, *18*(10), 1439–1463.

Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. *Perception & Psychophysics*, *63*(2), 272.

Hutchinson, J. M., Wilke, A., & Todd, P. M. (2008). Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches? *Animal Behaviour*, *75*(4), 1331–1349.

Lau, J. S., & Huang, L. (2010). The prevalence effect is determined by past experience, not future prospects. *Vision Research*, *50*, 1469–1474.

McNamara, J. (1982). Optimal patch use in a stochastic environment. *Theoretical Population Biology*, *21*(2), 269–288.

Nakayama, K., & Martini, P. (2010). Situating visual search. *Vision Research*,

Olsson, O., & S. Brown, J. (2006). The foraging benefits of information and the penalty of ignorance. *Oikos*, *112*(2), 260–273.

Pirolli, P. (2007). *Information foraging theory: adaptive interaction with information.* New York, NY: Oxford University Press US.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging Theory*. Princeton, NJ: Princeton University Press.

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.