

**Looking ahead: When do you find the next item in foraging visual search?**

**Anna Kosovicheva<sup>1\*</sup>, Abla Alaoui-Soce<sup>2,3</sup>, Jeremy M. Wolfe<sup>2,4</sup>**

This is the final copy of the following article, accepted for publication (currently in press):

**Kosovicheva, A., Alaoui-Soce, A., & Wolfe, J. M. (In press). Looking ahead: When do you find the next item in foraging visual search?. *Journal of Vision* .**

<sup>1</sup>Department of Psychology, Northeastern University

<sup>2</sup>Brigham and Women's Hospital

<sup>3</sup>Department of Psychology, Princeton University

<sup>4</sup>Harvard Medical School

\* Corresponding author:

125 Nightingale Hall

360 Huntington Ave.

Boston, MA 02115

E-mail: akosov@northeastern.edu

## Abstract

Many real-world visual tasks involve searching for multiple instances of a target (e.g., picking ripe berries). What strategies do observers use when collecting items in this type of search? Do they wait to finish collecting the current item before starting to look for the next target, or do they search ahead for future targets? We utilized behavioral and eye tracking measures to distinguish between these two possibilities in foraging search. Experiment 1 used a color wheel technique in which observers searched for T shapes among L shapes while all items independently cycled through a set of colors. Trials were abruptly terminated, and observers reported both the color and location of the next target that they intended to click. Using observers' color reports to infer target-finding times, we demonstrate that observers found the next item before the time of the click on the current target. We validated these results in Experiment 2 by recording fixation locations around the time of each click. Experiment 3 utilized a different procedure, in which all items were intermittently occluded during the trial. We then calculated a distribution of when targets were visible around the time of each click, allowing us to infer when they were most likely found. In a fourth and final experiment, observers indicated the locations of multiple future targets after the search was abruptly terminated. Together, our results provide converging evidence to demonstrate that observers can find the next target before collecting the current target and can typically forage 1-2 items ahead.

The ability to search for and find relevant items in a cluttered environment is crucial for performing a variety of daily functions, from finding your friend at the airport to finding a pen on your desk. Visual search has been studied extensively over the last several decades, most often in laboratory tasks that require observers to search for a single target that may or may not be present in an array of distractors. A large body of literature has examined the factors that influence how quickly and accurately observers can find individual items in these types of searches (Eckstein, 2011; Nakayama & Martini, 2011; Wolfe, 2018). In contrast, comparatively less is known about how visual search operates when a display contains a multiple, unknown number of instances of a target. Multiple-target search is characteristic of many real-world tasks, such as separating the quarters out of a pile of coins, picking the ripe berries from a bush, or searching for tumors in an X-ray.

A growing body of literature indicates that these types of searches are distinct from single-target search, providing important insights into search behavior in the real world. What strategies do observers use when collecting multiple targets? Previous work has investigated the rules that guide decisions to terminate search for additional targets, and to move to a new visual display (e.g., Cain, Vul, Clark, & Mitroff, 2012; Ehinger & Wolfe, 2016; Wolfe, 2013; Zhang, Gong, Fournie, & Wolfe, 2015). These decisions are akin to “patch leaving” decisions in foraging, as in when do you decide to stop collecting berries from the current bush and move to the next patch? Consistent with optimal foraging theory (Charnov, 1976), this work has shown that observers generally move to a new display when the instantaneous collection rate falls below the average rate for the environment.

At a more granular level, another line of research has investigated the target-to-target foraging strategies utilized *within* a single visual display or patch. This work has demonstrated that, when searching for multiple types of targets, observers’ strategies depend on the type of target. Observers switch frequently between target types when they are defined by a single feature, but organize their behavior into long ‘runs’, or sequential collection of the same item, when targets are defined by a conjunction of features (Á. Kristjánsson, Jóhannesson, & Thornton, 2014; T. Kristjánsson, Thornton, Chetverikov, & Kristjánsson, 2020). Furthermore, observers’ behavior within a single display can be partitioned into distinct phases, corresponding to different attentional operations (T. Kristjánsson et al., 2020). While these experiments generally speak to the efficiency of foraging search, observers are also prone to missing targets that would otherwise be found in single-target displays (subsequent search misses; e.g., Berbaum, 2012; Fleck, Samei, & Mitroff, 2010; Tuddenham, 1962). Together, this body of work has demonstrated that multiple target searches are distinct from searches for a single target, both in the high-level strategies that observers use, and in how visual attention is deployed within a given display.

While these studies have provided valuable insights into how attention is allocated in many real-world search tasks, much less is known about how foraging is coordinated with action (i.e., target collection). Specifically, when collecting items in a search display, do observers wait until they complete the current search (e.g., collect one berry) before beginning to look for the next item? Or can they plan ahead, searching for future targets before collecting the current item? And if they can search ahead, what is the limit or capacity for planning future target collection? The literature on visual working memory capacity would suggest that, in principle, observers should be able to accurately store the remembered locations of multiple future targets (for reviews, see Brady, Konkle, & Alvarez, 2011; Luck & Vogel, 2013). However, in a foraging task, observers must do more than simply remember the locations of future targets. To successfully plan ahead, the observer must perform several operations simultaneously: search for future targets, collect previously found items (i.e., implement some motor response), and keep a plan or a trajectory of future collection in working memory. Can these different operations be performed at the same time?

Previous work indicates that observers have at least some ability to perform these tasks in parallel and plan ahead when searching for multiple items. For example, Thornton and Horowitz (2004) instructed observers to search for and collect (i.e., click) multiple targets in a display where the remaining items were shuffled as soon as each target was collected. Compared to a condition in which the items remained in place, observers were considerably slower, as shown by an increase in the time between successive clicks. While this indicates that participants must be acquiring some information about future targets in multiple-target searches, a more complete picture of how observers coordinate target collection with visual search requires information about the timing of the relevant events. Specifically, when is the next item found, relative to the time that the current item is collected?

A challenge in determining the time at which observers find the next item is that we do not often know where observers are attending while they search for multiple targets. Although we can measure how long it takes observers to find a single target in a display, or measure the time interval between successive clicks in a display containing multiple targets (e.g., Horowitz & Thornton, 2008; Thornton & Horowitz, 2004), we would need to know where observers are attending at any given time point, to establish the time when the next item is found. In the present study, we adopted a multifaceted approach, in which we utilized several different methods for measuring the target-to-target time course of foraging search within individual displays. To briefly summarize these methods, the first experiment used a dynamic color wheel technique, in which observers foraged for T shapes among L shapes while all items changed color, providing an independent temporal reference or ‘clock’ for measuring finding times. At a pseudo-random moment, the search was abruptly terminated and observers were asked to indicate the

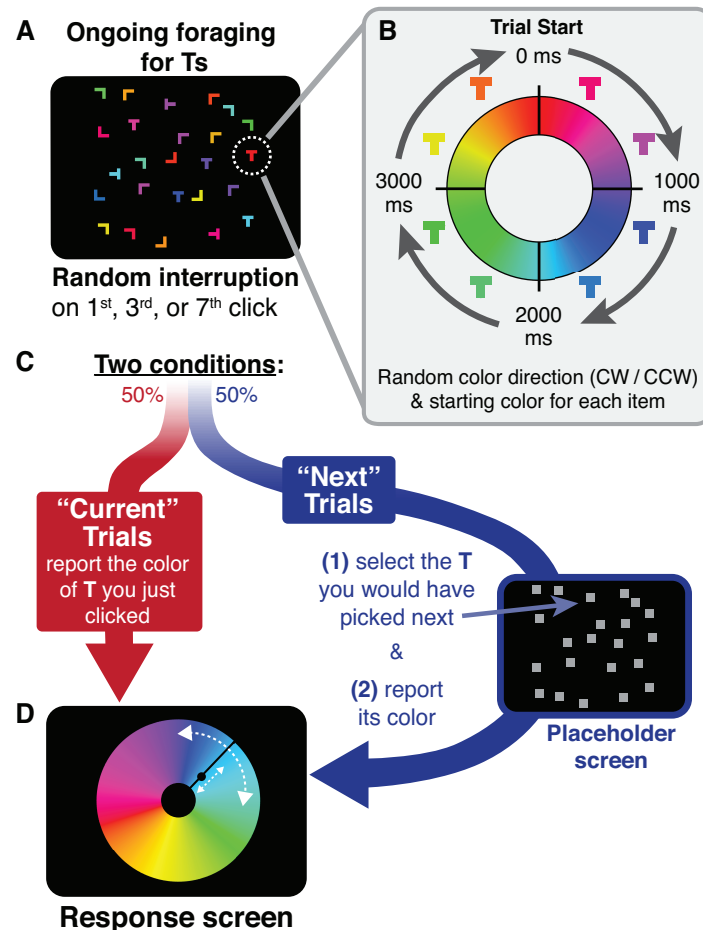
color of the item they were going to select next, allowing us to infer the time that the item was found. We validated these responses in a second experiment, in which we recorded eye movements while observers performed the same task. Fixation onset times towards the end of the trial were consistent with observers' color responses, and more broadly, fixation behavior was consistent with a searching-ahead strategy throughout the trial. In a third experiment, we developed a novel method in which items were intermittently occluded. We inferred target-finding times from the history of item visibility, based on the logic that items had to have been found at times they were visible. Taken together, our results from these different methods provide converging evidence that observers begin to search for the next item before collecting the current target. In a fourth and final experiment, we tested how many items observers can plan ahead to collect. In other words, what are the capacity limits for foraging ahead?

### **Experiment 1: Color Self-Reports**

In Experiment 1, we used observers' self-reports to estimate the moment in time when targets are accessed during foraging. Previous studies examining other perceptual and cognitive processes have used such self-reports, based on an independent temporal reference, to measure the time elapsed while an observer performs a task. An early variant of this type of method was developed by Wilhelm Wundt (1883), who used a clock to measure the delay between the time that a sound is played and the time that it is experienced. More recent studies have used videos of a changing clock to measure shifts of attention (Carlson, Hogendoorn, & Verstraten, 2006), as well as observers' perceived gaze direction around the time of a saccade (Hunt & Cavanagh, 2009). A similar technique can be applied to other visual tasks, involving systematic changes along other feature dimensions, such as changes in the color of text while reading (Kosovicheva & Bex, 2016), or the color of targets to track shifts of attention (Callahan-Flintoft & Wyble, 2017). In these experiments, a target (e.g., the text of a sentence) gradually cycles through a set of hues over time (e.g., from red to yellow to green), and observers use a color response palette to report the color of an item when they attended to it.

Together, these studies have demonstrated that observers can reliably report the time that items are processed or attended based on an independently changing feature dimension. We therefore adapted this method to measure the dynamics of visual foraging search with color-changing targets. By systematically and independently varying the colors of items in the display over time, we introduced a temporal reference that would minimally interfere with a foraging search task based on item shape. In the experiment, observers searched for 2-16 T shapes among L shape distractors in a 25-item display (Figure 1A) while all items continuously varied independently in hue from 0 to 360° around a circle of hues. Each item started with a random color phase and direction of color change (Figure 1B). The trial

terminated unpredictably after a pseudorandom number of targets had been clicked (1, 3, or 7). At that point, observers were shown a color response palette and asked to report, on separate trials, either the color of the “Current” target they had just clicked or the “Next” target they intended to click (Figures 1C and 1D). Observers reported the location of their intended “Next” target on a placeholder screen consisting of boxes covering each item. Observers did not know in advance when the trial would terminate or what trial condition (Current or Next) they would complete. We then used the difference between the observer’s color response and the actual color of the item at the end of the trial to infer the time, relative to the end of the trial, that they found the target. Observers’ responses indicated that they could reliably report both the locations and colors of “Next” items above chance levels. We used observers’ color reports to test the prediction that color reports for Current targets would be earlier than color reports for Next targets. In addition, we varied the number of targets remaining at the end of the trial, to test the prediction that as fewer items remain in the display, observers would take longer to find the scarcer, Next item, and report colors that were presented later in the trial.



**Figure 1. Procedure for Experiment 1.** (a) Observers continuously searched for and collected 2-16 T shapes among 9-23 L shape distractors, while (b) each item in the display continuously cycled through a

360° range of hues. Each item had a randomly chosen direction (clockwise or counterclockwise) and phase. (c) The display was abruptly terminated on the 1<sup>st</sup>, 3<sup>rd</sup> or 7<sup>th</sup> click, at which time, observers made responses in one of two conditions. In Current Trials, observers reported the color of the T that they had just clicked. In Next Trials, observers were first shown a placeholder screen, in which all items were occluded by gray squares. Observers indicated both the location of the T they would have clicked next, and its color. (d) Observers adjusted a cursor on a color response palette to indicate their responses along two axes, where the angle of the black line corresponded to hue, and the distance of the response bead (black dot) from the center indicated their confidence in their color response.

## **Methods**

### **Observers**

Twelve observers (9 female, mean age: 28.25, range 20 – 39) participated in Experiment 1. Power calculations, based on effect sizes observed with a comparable color report technique (Kosovicheva & Bex, 2016) indicated that a minimum of 8 observers was necessary to detect a difference between the trial conditions at 95% power (Cohen's  $f = 1.13$ ;  $\alpha = .05$ ). All observers had normal or corrected-to-normal vision and passed the Ishihara Color Test. Observers gave informed consent prior to participating in the experiment. Procedures were approved by the Institutional Review Board at Brigham and Women's Hospital and followed the tenets of the Declaration of Helsinki. All observers were naïve to the purpose of the experiment.

### **Stimuli**

The experiment was run on an iMac computer (Apple, Inc., Cupertino, CA) with a 24" LCD panel and a screen resolution of  $1920 \times 1200$  pixels (60 Hz refresh rate). It was programmed in Matlab (The MathWorks, Inc., Natick, MA) using the Psychophysics Toolbox Version 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Observers viewed the display binocularly at a distance of approximately 57 cm, with the display region subtending 48° horizontally and 31° vertically.

Stimuli were shown on a uniform black background. Each search display consisted of 25 items, with 2, 4, 8, or 16 'T' shapes (targets) among 'L' shapes (distractors), randomly intermixed within a 5 row  $\times$  5 column screen-centered grid. Non-overlapping item locations on each trial were assigned by uniformly distributing the item centers across an area 12.2° wide  $\times$  12.2° high, and then adding a random spatial jitter to each item between -1.12° and +1.12° independently in both the horizontal and vertical directions. Each shape was 0.8° in height and width (line width of 0.24°) and assigned a random spatial orientation of 0°, 90°, 180°, or 270°. Item colors were manipulated in HSV color space, varying in hue from 0 to 360°, while saturation and value were both maintained at 100%. Each item was randomly assigned an initial hue between 0 and 360° and cycled continuously through the full 360° range of hues at a uniform rate of 0.25 cycles per second (increment of 1.5° per frame). To minimize the possibility of

observers anticipating or monitoring the direction of color change, all items in the display were independently assigned a random color direction (clockwise vs. counterclockwise). No particular effort was made to precisely control luminance or other aspects of color since the colors are merely used as a clock and are not, themselves, critical to the task.

## **Procedure**

Observers were instructed to collect the T shapes as quickly as possible by clicking on them with a mouse cursor, and auditory feedback was provided to encourage observers to maintain a fast collection rate. If the time from the previous click on a T shape (or from the beginning of the trial) exceeded 1250 ms, observers heard a pure tone (800 Hz, 300 ms duration) at regular 700-ms intervals until a T shape was clicked. Each T shape was removed from the screen as it was clicked, and the trial randomly terminated after the 1<sup>st</sup>, 3<sup>rd</sup>, or 7<sup>th</sup> T was collected.

Observers completed two types of trials (“Current” and “Next” trials), which were randomly interleaved, and identical up until the time of the interruption. Observers did not know when the trial would terminate or which condition they were completing until after the click that ended the trial. On Current trials, observers were instructed to report the color of T that they had just clicked. On Next trials, all remaining items in the display (both Ts and Ls) were immediately replaced by gray boxes, and observers were instructed to: (1) click on the box corresponding to the item that they would have picked next, and (2) report its color.

For both types of trials, observers reported their color selection on a response wheel, which was shown immediately after the final click (Current trials), or after the gray box was selected (Next trials). The response wheel was a screen-centered annulus (inner and outer radii of 1° and 13.7°, respectively) that varied in hue from 0° to 360° as a function of spatial angle in equal steps. Observers used the mouse to control an angular cursor (a black line, 12.7° long and 0.08° wide) positioned inside the annulus to indicate their color selection. In both types of trials, observers were also instructed to report their level of confidence in their color report by changing the position of a circular bead (0.43° diameter) attached to the cursor. Observers could change the distance of the bead from the center of the wheel, where the inner and outer radii corresponded to least and most confident, respectively. Observers therefore selected their response along two axes, where angle corresponded to hue, and distance from the center corresponded to confidence, and both responses were recorded with a single click. The initial angle of the cursor and the orientation of the response wheel were randomized on each trial. After making their response, observers were shown a feedback screen listing the number of targets collected, the total trial



duration, and the average time per item on the previous trial. Observers then clicked the mouse button when ready to continue to the next trial.

The combination of the initial number of targets (2, 4, 8, and 16) and the number of clicks on which the trial terminated (1, 3, and 7) resulted in 9 possible Target Densities at the end of the trial (Density = # targets / total # of items) Density ranged from 4.17% to 62.5%. There were 28 trials for each unique combination of Trial Type (Current vs. Next) and Target Density, which were randomly interleaved. Observers completed a total of  $28 \times 2 \times 9 = 504$  trials, divided into 2 blocks of 252 trials each. At the beginning of the experiment, observers completed a short practice block of 20 trials, consisting of 5 Current Trials, followed by 5 Next Trials, and then 10 trials with the two Trial Types randomly interleaved.

## **Data Analysis**

For each trial, we calculated the difference between the observer's color report (0-360° in hue), and the actual color of the corresponding (Current or Next) item at the end of the trial. Since items cycled through the hues at a constant rate, we converted this color difference to a time difference, represented in milliseconds. Color responses were then averaged across trials. Because colors are identical at -2000 ms and +2000 ms relative to the time of the click, these averages took into account the circular wraparound in the data. Circular statistics (e.g., mean, standard deviation) were calculated using CircStat toolbox for Matlab (Berens, 2009). Color report times (-2000 through +2000 ms) were scaled to represent angles from 0 to  $2\pi$  radians prior to calculating the average for each observer and each condition, and then converted back to milliseconds.

In addition, to test whether observers were guessing randomly on the color report task, we compared the width of the distribution to that expected by random guessing. This was done with a permutation test procedure in which we shuffled the mapping between responses and the colors shown for each observer, such that differences were calculated between the observer's color report on one trial, and the actual color on a different trial. We then calculated the standard deviation of the response distribution, averaged across Current and Next trials. This was repeated for 1000 iterations to generate a permuted null distribution of the expected response SDs expected by random guesses, and we calculated the proportion of iterations in which the SD in the null distribution was less than the observed SD.

## **Results**

### **Identification on placeholder screen**

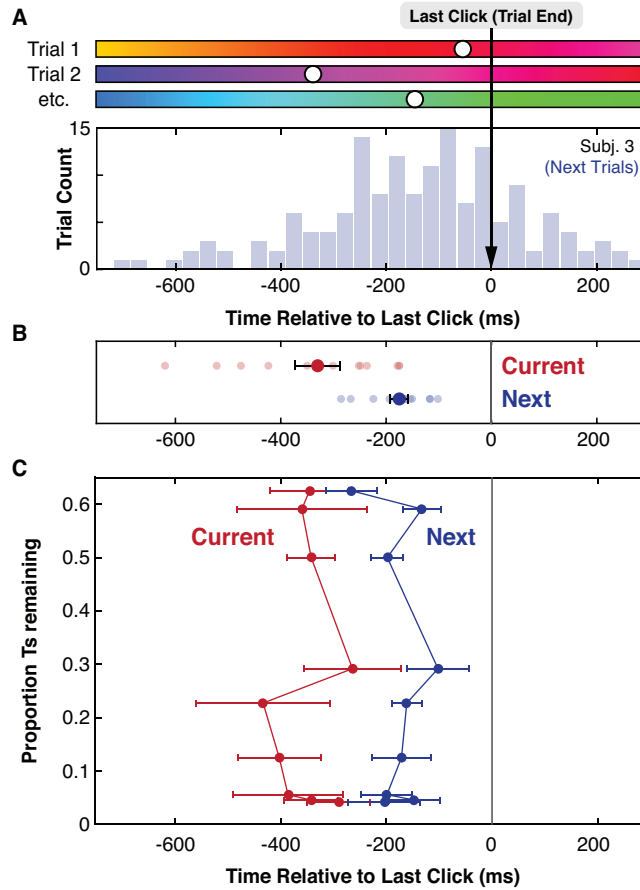
We first analyzed the “Next” trials to determine whether observers could correctly identify Ts on the placeholder screen. In other words, do observers select Ts when they’re asked to identify the target they intended to click on next? The expected chance level of performance was calculated by averaging, across trials, the proportion of targets remaining at the end of the trial, out of the total number of items remaining in the display at the end of the trial. A one-sample t-test indicated observers performed significantly better than the chance performance rate (66.4% vs. 27.8%),  $t(11) = 13.7$ ,  $p < .001$ , and were consistently above chance across all Target Density conditions (Figure S1). Thus, at the instant that they clicked on the Current T and the trial unpredictably ended, observers could usually identify the Next T. As inaccurate trials would contribute additional variability to the color reports, only correct trials were used for the remaining analyses in the Next condition (see Figure S2 and Supplemental Note 1 for analysis with these trials included). All trials were analyzed in the Current condition.

### Color reports

We next analyzed observers’ color reports, comparing the mean and standard deviation between Current and Next trials. Figure 2A shows a sample distribution of times corresponding to reported colors for one observer. If observers were randomly guessing when reporting the colors of targets, these responses would be uniformly randomly distributed. In contrast, results of the permutation test showed that the SD of the response distribution was narrower than that of the permuted null distribution for each of the twelve observers (all  $p$ -values  $< .001$ , which were compared to a Bonferroni-corrected alpha,  $\alpha_B$ , of .004). In addition, there was no significant difference in the SD of the color reports between Current and Next trials (573 and 526 ms, respectively,  $t(11) = 1.14$ ,  $p = 0.28$ ). To verify that inaccurate identification of target location on Next trials also produced random color responses, we calculated the SD of the response distribution, analyzing only the Next trials that were mislocalized. The observed SD here was much closer to the average SD from the permuted null distribution (811 vs 855 ms, respectively) and not significantly lower than the null distribution for seven of the twelve observers ( $p$ -values  $> .07$ ).

Figures 2B shows the mean values of the color report distributions, separately for the Current and Next trials. A 2 (Trial Type: Current vs Next) x 9 (Target Density) repeated-measures ANOVA on reported time showed a significant main effect of Trial Type,  $F(1,11) = 12.97$ ,  $p = .004$ ,  $\eta_p^2 = .54$ . On average, observers reported earlier colors (relative to the end of the trial) in “Current” trials compared to “Next” trials (-330 ms and -175 ms, respectively). Figure 2C shows the mean reported times, separated by Target Density condition (proportion of targets remaining at the end of the trial). There was no main

effect of Target Density,  $F(8,88) = .76$ ,  $p = .64$ ,  $\eta_p^2 = .06$ , or interaction between Target Density and Trial Type,  $F(8,88) = .53$ ,  $p = .83$ ,  $\eta_p^2 = .05$ .



**Figure 2. Analysis and results for Experiment 1.** (a) The difference between observers' color responses (white circles) and the color of the item at the end of the trial (black arrow) was used to calculate the time corresponding to the selected color, relative to the last click (0 ms). This was calculated for each trial to generate a distribution of times corresponding to color responses (histogram for one representative observer), separately for the Current and Next conditions. (b) Mean time corresponding to color reports, relative to the last click, for the Current and Next conditions (red and blue, respectively). Light-colored scatter points represent individual observers. (c) Data in (b), replotted to show mean time of color report as a function of the proportion of Ts remaining on the screen. Error bars represent  $\pm 1$  standard error of the mean (SEM).

## Discussion

Experiment 1, using a color-report method to measure shifts of attention during foraging search, yielded two main findings. First, observers typically knew where they were going next and could reliably report information about the item, demonstrating an ability to search for additional targets while collecting previous targets. On Next trials, in which observers had to identify the location of the target that they intended to click next, targets were selected over distractors well above chance performance levels. Observers were also reasonably precise in their color reports; distributions for Current and Next

trials were narrower than those expected by random guesses, indicating that observers were also able to report information about the color of the Next item. This is also consistent with previous results demonstrating that observers can report information from an independently changing feature dimension while performing an unrelated task (e.g., Kosovicheva & Bex, 2016). Second, the color report results also provide information about the times that observers attend to items in the display, with observers reporting colors corresponding to 330 ms before the last click on Current trials, and 175 ms before the last click on Next trials. These color reports times raise the question of where observers were attending throughout the trial. The average interval between clicks was 1025 ms per item, yet the color reports between consecutively clicked items were only 155 ms apart. This leaves a large portion of the trial unaccounted for: what are observers attending to during the remaining interval in between clicks? It is likely that observers are continuously visiting items in the display (Ts and Ls), and report the colors of the items when they *most recently* attended to them. From these results, we can infer that for Current targets, observers typically committed to the click approximately 330 ms before the click. For Next targets, observers attended to that item between that commitment to the Current target and the click on that Current target. On average, observers were attending to the T that they would collect next, 175 ms before they collected the current T. Together, these results demonstrate that observers actively search for and find the Next target before they finish collecting the Current target.

Surprisingly, we observed no relationship between the number of items remaining and the time associated with the reported color. As fewer targets remain in the display, we would expect observers to take longer to find Next items, and therefore would expect color reports to shift to later times when there are few targets remaining. However, consistent with this hypothesis, we observed a decrease in accuracy on the placeholder screen as fewer targets remained in the display (Figure S1). The absence of an effect on color reports might be due to the reduced number of trials in the analysis when fewer targets remain in the display (28% of trials in the lowest Target Density condition had correct identification responses on the placeholder screen). Nevertheless, observers' ability to reliably report information about the Next target indicates that they frequently find Next items prior to the click on the Current item.

One concern is the possibility that performance in this color wheel task may not be representative of observers' strategies in a more typical foraging task. It is possible that observers were deliberately adjusting their strategy in the color wheel task because they knew that they would have to report information about future targets on half the trials. To test this idea, we varied the relative frequency of Current and Next trials in a separate experiment. Observers completed two separate blocks of trials, one with 25% Next and 75% Current trials, and another with an even (50-50%) distribution between the two conditions, as in Experiment 1. If observers are changing their strategy based on the

task demands, we would expect color responses to vary between these two conditions. However, results of this experiment indicated that neither the precision of observers' color reports, nor the mean reported times were affected by this change in relative condition frequency (see Supplemental Note 2).

While these results give us with a picture of when observers are attending to different items in the display, we sought to validate these results with other measures. We therefore explored observers' foraging strategies further by recording observers' gaze positions while performing the same task. This also allowed us to compare observers' color reports to the times that observers fixated the targets, and to measure fixation behavior throughout the trial to determine the time when items are initially fixated.

### **Experiment 2: Fixations during foraging**

In Experiment 2, we recorded observers' eye movements while they performed the foraging task from Experiment 1 in order to determine whether observers' color reports in Experiment 1 generally agree with recorded fixation behavior. That is, do observers fixate the Next clicked item about 175 ms prior to the end of the trial, as predicted by their color reports? An advantage of eye tracking over self-report methods is that it also allows us to measure fixation behavior throughout the entire trial, rather than assessing only when the last item in the trial was attended. However, this comes with a loss of temporal resolution, as observers typically make 2-4 fixations per second (in contrast to the continuous nature of the color reports), and these fixation locations may not necessarily reflect where observers are attending at any point in time. Nevertheless, eye tracking gives us the opportunity to acquire converging evidence for the conclusions of Experiment 1.

We ran two experiments with different groups of observers. Experiment 2a was similar to Experiment 1, with the addition of gaze recording while performing the task, allowing for a comparison between color self-reports and gaze position measures. As mentioned previously, one limitation of the color-report method is that observers may change or adapt their normal strategy to meet the demands of this task. Therefore, we conducted a second experiment (Experiment 2b), to evaluate eye movements in the absence of the demands imposed by the color report task. These observers had not previously completed the color-report task, were given no specific instructions other than to collect the targets. In 2b observers did not report the colors of the items. The two experiments were otherwise as similar as possible, and items cycled through the color wheel in both experiments.

## **Methods**

### **Observers**

Thirteen observers, including one author, participated in Experiment 2a, and thirteen different observers participated in Experiment 2b. One observer in Experiment 2a did not complete the experiment due to an inability to calibrate the eye tracker, and one observer in Experiment 2b was removed from the analysis due to poor calibration in the eye tracking procedure (mean error on validation of 1.5°). The final sample consisted of 12 observers in each experiment (respectively, 8 and 7 female participants; mean ages 28.7 and 30.3, ranges 19-42 and 18-48).

## **Eye Tracking**

Eye movements were recorded using an Eyelink 1000 desktop infrared eye tracker (SR Research Ltd., Mississauga, Ontario, Canada), used in conjunction with the Eyelink Toolbox for Matlab (Cornelissen, Peters, & Palmer, 2002). Observers viewed the display binocularly, and the right eye gaze position was recorded for each observer at a sampling rate of 1000 Hz. Observers completed a standard 9-point calibration procedure (Stampe, 1993) prior to each block of trials (mean error on validation of 0.54° and 0.57° for Experiments 2a and 2b, respectively). Noise artifacts were reduced using Eyelink software, which applied a heuristic filtering algorithm to the raw gaze position samples (see Stampe, 1993, for details). Gaze information was then parsed into saccades and fixations, with velocity and acceleration thresholds of 30°/s and 8000°/s<sup>2</sup>, respectively.

## **Stimuli and Procedure**

Stimuli and procedure were identical to Experiment 1, with the following exceptions. The experiment was run on a PC with a Mitsubishi Diamond Pro 91TXM CRT monitor subtending 38° × 29° at a viewing distance of 57 cm. Screen resolution was set to 1024 × 768 pixels, and the refresh rate to 75 Hz. The display was scaled so that item locations on each trial were assigned by uniformly distributing the 5 × 5 array across an area 17.6° wide × 15.6° high, and then adding a random spatial jitter of up to ±1.62° horizontally and ±1.39° vertically. Items were 1.15° in height and width (0.34° line width), and colors cycled through the full 360° set of hues at a rate of 0.31 cycles per second (1.5° per frame at 75 Hz).

In Experiment 2a, observers responded as in Experiment 1, reporting the color of either the item they had just clicked on (Current Trials) or the item they would have picked next (Next Trials), as well as their confidence in the color report. In Experiment 2b, observers were shown the same color-changing search displays, but were not instructed to report any color. Instead, the experiment automatically advanced to the feedback screen (as in Experiment 1, showing the number of targets collected and time per item). Search displays contained 4, 6, 10, or 14 Ts, and trials terminated after the 3<sup>rd</sup>, 5<sup>th</sup>, or 7<sup>th</sup> T

collected, resulting in 9 Target Density conditions (between 4.5% and 50% Ts). The increase in the initial proportion of Ts in the display produced slightly faster collection rates in Experiment 2a than in Experiment 1 (mean interval between clicks of 754 vs 1025 ms, including the interval between display onset and first click). In Experiment 2a, observers completed 12 trials for each possible combination of Target Density and Trial Type, for a total of 216 trials. In Experiment 2b, observers completed 48 trials for each of the Target Density conditions, for a total of 432 trials (as there were no color reports, the Trial Type conditions were omitted, and the total number of trials was increased). After a short practice block, observers completed four blocks of trials, lasting approximately 10-15 minutes each, and observers were re-calibrated at the beginning of each block.

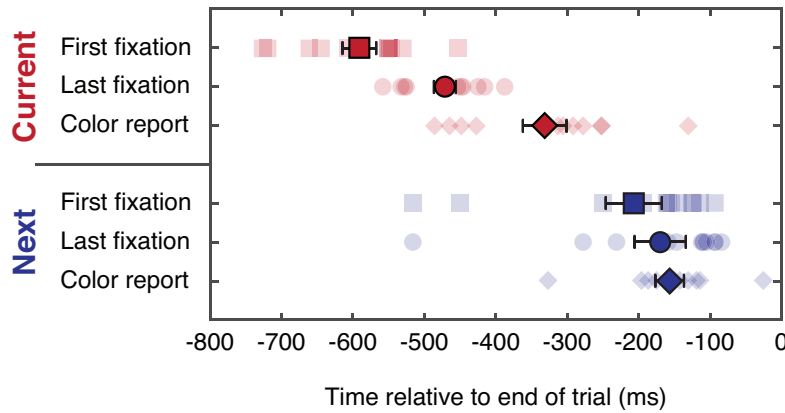
## Data Analysis

Color reports were analyzed using the same procedures as in Experiment 1. We performed three different analyses on the gaze data. In each analysis, fixations were considered to fall on the item (T or L) if they were within a  $2.15 \times 2.15^\circ$  square region centered on the item. Reported fixation times are the times of fixation onset. First, for Experiment 2a, we determined time associated with the observers' color report as well as the time of the first and last fixations on the corresponding target, separately for Current or Next trials. To facilitate direct comparisons to the color reports, in the Next condition, we only analyzed the trials in which observers correctly identified Ts on the placeholder screen. In trials in which there was only one fixation on the corresponding item, the same fixation time was used for both (first and last fixation) analyses. Second, for both experiments, we extracted the time points of three different events relative to each click for each item (referred to as item N), across all clicks and across all stimulus conditions: the onset of the first fixation on that item ( $\text{Fixation}_N$ ), the onset of the first fixation on the next item ( $\text{Fixation}_{N+1}$ ), and the time of the click on the next item ( $\text{Click}_{N+1}$ ). Finally, for both experiments, we determined where observers were fixating, at different time points near the time of each click, using a region of interest (ROI) based analysis. For each click, we extracted the fixation coordinates (blinks and saccades excluded) for a set of time points  $\pm 1300$  ms on either side of the click, in 20 ms increments. This location was then assigned to one of four categories:  $\text{Target}_N$  (the current clicked item),  $\text{Target}_{N+1}$  (the next clicked item), Other Target, and Any L. The fixation categories at each time point were then used to calculate a density function showing the proportion of fixations at each time point, in each bin, across all clicks. Fixations outside of these four regions (i.e., on blank areas of the screen) were not assigned to a category, but included in the total fixation count (i.e., the denominator) when calculating the proportion of total fixations.

## Results

### Color reports

Color reports were collected only in Experiment 2a, and were similar to the results in Experiment 1. Briefly, observers generally correctly identified Ts on the placeholder screen above chance levels (69.3% of trials; chance performance rate of 25.6%,  $t(11) = 11.24$ ,  $p < .001$ ), and averaged across conditions, the width of the color response distributions were narrower than those expected by chance (permutation test; for all observers,  $p$ -values  $< .001$ ). In addition, observers reported earlier colors for Current trials (-332 ms relative to click time) compared to Next Trials (-157 ms),  $t(11) = 5.36$ ,  $p < .001$  (Figure 3; diamond symbols).



**Figure 3. Color reports and corresponding fixation times in Experiment 2a.** End-of-trial fixation times for the first and last fixations (squares and circles, respectively) and times corresponding to color reports (diamond symbols) in Experiment 2a. Current and Next targets are shown in red and blue, respectively. Error bars represent  $\pm 1$  SEM.

### End-of-trial fixations: Comparison to color reports

For Experiment 2a, in which observers reported the colors of Current and Next clicked items at the end of the trial, we analyzed fixation onset times for the corresponding items. We note that this is a somewhat limited analysis of the gaze data, since it only includes fixation onset times towards the end of the trial, rather than all fixations throughout the trial. The Current clicked item was fixated at least once on 82.7% of all trials (mean number of fixations: 1.18), and the item identified as the Next item was fixated at least once on 39.7% of all trials (mean number of fixations: 0.47). Figure 3 shows the onset times for the first and last fixations on the items for which the colors were reported. For Current targets, first and last fixations were 591 and 471 ms before the end of the trial, respectively. For Next targets, we analyzed only trials with correct identification responses; for these trials, first and last fixations were 207 and 170 ms before the end of the trial. In each case, observers reported colors that occurred slightly after



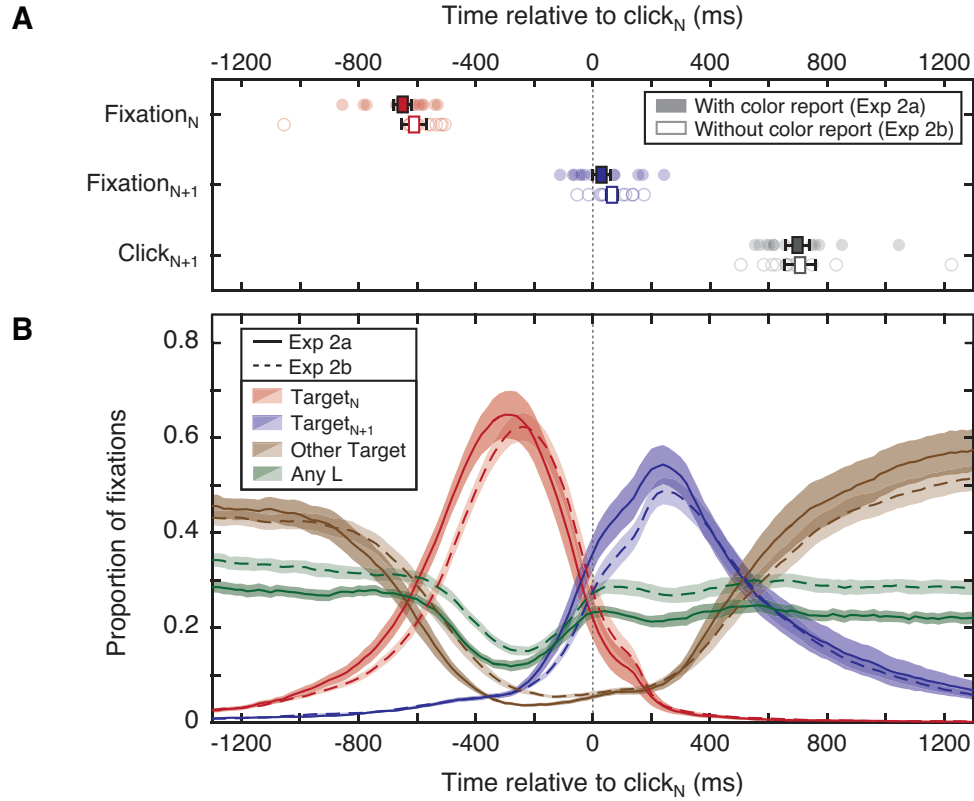
fixation onset. Color reports lagged behind fixation onsets by 139-259 ms for Current items, and by 13-50 ms for Next items. The fixation onset times and color responses shown in Figure 3 were analyzed together with a 2 (Trial Type: Current or Next) x 3 (Event Type: First Fixation, Last Fixation, Reported Color) repeated-measures ANOVA. Event onset times showed a significant main effect of Trial Type  $F(1,11) = 81.23, p < .001, \eta_p^2 = .88$ , with color reports and fixation times occurring earlier for Current trials, relative to Next trials. In addition, there was both a significant main effect of Event Type  $F(2,22) = 16.82, p < .001, \eta_p^2 = .60$ , and a significant interaction between Trial Type and Event Type,  $F(2,22) = 25.07, p < .001, \eta_p^2 = .70$ . Post-hoc tests for main effects of Event Type at each level of Trial Type showed a significant main effect of Event Type for Current targets,  $F(2,34.47) = 36.11, p < .001$ , but not for Next targets,  $F(2,34.47) = 1.45, p = .25$ , indicating that the small differences across event types (i.e., between fixations and color reports) for Next items were not significant. Together, the results in Figure 3 demonstrate that the color reports correspond approximately to the times that these items were last fixated, although for Current targets, they lag behind the time of fixation.

We note that one constraint of this analysis—for both color reports and fixation onsets—is that the mean values are necessarily less than 0 ms (the end of the trial), as the trial ends abruptly on the last click, and we only analyzed fixations that occurred while the search was ongoing. Although responses on individual trials in the color task may be later than 0 ms due to random variability in color responses, we would expect the mean values to be negative, as observers are given no information about the colors or identities of the items after that point. However, it is possible that throughout each trial, in the process of collecting Ts, observers sometimes fixate the next clicked item *after* the click on the current item. Therefore, we also analyzed gaze position not just at the end of the trial, but throughout the trial as well, to determine when the “next” items were typically fixated.

### **All fixations relative to time of click**

To determine when items were typically fixated relative to each click, we analyzed the fixation onset times for every click across all trials (without regard to trial type condition, i.e., Current vs Next trials), separately for Experiments 2a and 2b. For every clicked item (referred to as item  $N$ ), we calculated the mean onset time of the first fixation for the corresponding item (Fixation <sub>$N$</sub> ), as well as the mean onset time of the first fixation on the next clicked item (Fixation <sub>$N+1$</sub> ). In addition, we calculated the time of the click of the next item (Click <sub>$N+1$</sub> ). All three values are reported relative to the time of the click on the current item (Click <sub>$N$</sub> ). Unlike the previous analysis, the clicks occur mid-trial, and therefore the mean values for the fixation on the next clicked item are not restricted to negative values. Figure 4A shows times of all three events relative to the time of Click <sub>$N$</sub> . Averaged across both experiments, the mean onset times for Fixation <sub>$N$</sub>  and Fixation <sub>$N+1$</sub>  were, respectively, -630 ms and +48 ms. Notably,

compared to the results in Figure 3, fixation times for the next clicked item shift to positive values because some of the fixations occurred after the click on the current item. On average, Click<sub>N+1</sub> occurred 703 ms after the previous click (not including the interval between display onset and first click).



**Figure 4.** Fixation results from Experiments 2a and 2b, analyzed from all fixations and all clicks. (a) Time of fixation onset on each item (Fixation<sub>N</sub>), the time of fixation onset on the next clicked item (Fixation<sub>N+1</sub>), and the time of the click on the next item (Click<sub>N+1</sub>) relative to the time of the click on the current item (Click<sub>N</sub>). Values are shown separately for data collected with and without the additional color report task (Experiments 2a and 2b, shown in filled and empty scatter points, respectively). (b) Across all clicks, the proportion of fixations at each time point in each of four regions, as a function of time relative to the click on the current item. Unlike panel (a), which shows fixation onsets, panel (b) shows the locations of fixations based on their x,y coordinates across the same set of time points (i.e., where observers are looking at any given point in time). Experiments 2a and 2b shown in solid and dashed lines, respectively. Error bars and filled regions represent  $\pm 1$  SEM.

Figure 4A shows the differences between the times of these three Event Types, as well as showing the rather inconsequential differences between the two experiments (presence or absence of the color task). These results show that onset of the first fixation on the target (N or N+1) occurs ~600-700 msec before it is collected. Moreover, it shows that the N+1 target is first fixated around the time that the Nth target is collected. The times were directly compared with a 3 (Event Type) x 2 (Color Task) mixed-model ANOVA. Event Type was a within-subjects factor, and the presence or absence of the

Color Task (i.e., Experiment 2a or 2b) was a between-subjects factor. Unsurprisingly, we observed a significant main effect of Event Type,  $F(2,44) = 516.59, p < .001, \eta_p^2 = .96$ , and post-hoc contrasts using Tukey's HSD tests showed significant differences in all pairwise comparisons between Event Types, (all  $p$ -values  $< .001$ ). Neither the main effect of Color Task:  $F(1,22) = 1.35, p = .26, \eta_p^2 = .06$ , nor the interaction between Color Task and Event Type,  $F(2,44) = .08, p = .92, \eta_p^2 = .004$  were significant. The absence of an effect of Color Task (Exp 2a vs. Exp 2b) indicates that the expectation of being asked about the next item at any point did not affect either observers' fixation behavior or their collection rates. In addition, we tested whether the time of Fixation<sub>N+1</sub> was significantly different from the time of the Click<sub>N</sub> (0 ms). In Experiment 2a, the mean fixation onset time on the next clicked item (30 ms) was not different from 0 ms (the time of the click),  $t(11) = .92, p = .38$ , but was significantly above zero in Experiment 2b (65 ms),  $t(11) = 3.34, p = .007$ .

Figure 4B shows the distribution of all fixations on individual items relative to the time of Click<sub>N</sub> using an ROI-based analysis, in which fixations were assigned to four categories based on their location: Target<sub>N</sub>, Target<sub>N+1</sub>, Other Target, or Any L. Unlike Figure 4A, which shows the time corresponding to the *onset* of each event, Figure 4B shows which items observers were fixating at any given moment, based on the  $x,y$  coordinates of their fixations at each time point. As expected, the distribution for each item (e.g., Target<sub>N</sub>) shows a steep increase at approximately the time of the average fixation onset for the corresponding item in Figure 4A (e.g., Fixation<sub>N</sub>). Note that the peaks of these distributions are shifted slightly later relative to the mean time of the onset of the first fixation (Figure 4A), partly due to the occurrence of re-fixations, and the fact that fixations typically last 200-300 ms (mean fixation durations were 286 ms and 279 ms for Experiments 2a and 2b, respectively). Figure 4B tells a fairly clear story. Observers fixate on targets several hundred milliseconds before they are collected. By the time that item N is clicked, the eyes have moved on to item N+1. By the time that N+1 is clicked, the eyes are off acquiring some other target. If observers were randomly fixating Ts and Ls, one would expect a higher rate of fixation on Ls. The relatively low rate of L-fixation suggests that the Ts are being detected before they are fixated. Thus, the estimates of when observers first find the Next target should be regarded as conservative. The T was probably found peripherally, then fixated, and eventually clicked.

Supplemental Movies S1-S3 provide a more dynamic look at the time course of these events. They show sample two-dimensional 'maps' of individual fixation onsets as well as the density of fixated locations frame-by-frame across the same set of time points, where the spatial locations of Target<sub>N</sub> and Target<sub>N+1</sub> were aligned across clicks.

## Discussion

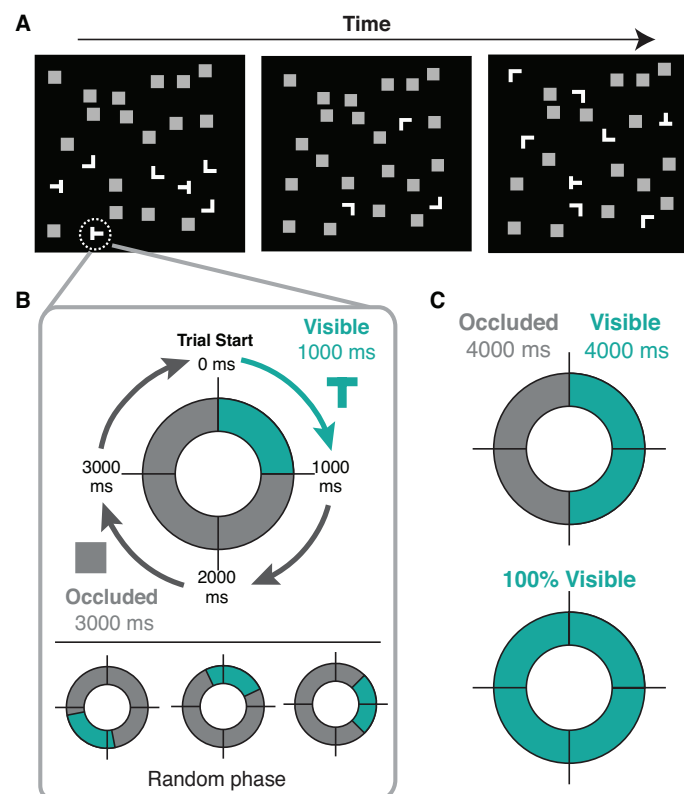
Experiment 2a replicated the pattern of results observed in Experiment 1, showing that observers could typically report both the color and the location of the next item that they intended to click. Color reports also showed that observers determined where they were going next sometime after committing to the click on the current target, but before the end of the trial. Moreover, Experiment 2a demonstrated that observers' color reports were consistent with the times that items were fixated. As shown in Figure 3, for Current items, observers typically reported colors that occurred 139-259 ms after they fixated them; for Next items, this delay was approximately 13-50 ms. However, fixations and color reports for Current and Next items both have the constraint that average values must come from times that are before the end of the trial (values less than 0 ms). Therefore, these values are likely limited by a ceiling effect, in which the mean values are compressed towards the last click. In contrast, when we analyzed *all* clicks that occurred throughout each trial, which do not have this constraint, we found that observers typically fixate the next clicked item close to the time of the current click (on average, 30 and 65 ms after the click for Experiments 2a and 2b, respectively). In some cases, these fixations occurred well before the time of the click on the next item. As shown in Figure 4B, the density of fixations on the next item begins to increase as early as -200 ms relative to the click, and peaks at approximately 200 ms after the click. Together, these results indicate that at least some of the time, then, observers must be finding the next item before they click on the current item.

In addition, in a direct comparison between Experiments 2a and 2b, we showed that the observed patterns of fixation behavior do not change when observers are asked about the next target, suggesting that the observed results in Experiment 1 reflect strategies in a typical visual foraging task, and are minimally influenced by task demands. In other words, observers plan at least one step ahead, regardless of whether they expect to be asked about the next item. Next, we sought to independently validate these results using a novel occlusion-based method for measuring the times that individual items were attended.

### **Experiment 3: Intermittent occlusion**

In Experiment 3, we examined whether the results observed in the first two experiments could be reproduced with a different method for measuring the time that individual items were attended throughout the course of the trial. Figure 5 outlines the general procedure, in which we systematically varied item visibility by intermittently occluding individual items in the display. As in Experiments 1 and 2, observers continuously foraged for Ts among Ls (see Figure 5A). The new aspect of Experiment 3 was that items were only visible intermittently. Both Ts and Ls were occluded by gray squares with a

regular on-off cycle (e.g., 1 second visible, 3 seconds occluded; Figure 5B). Each item had a randomly selected phase, limiting the observers' ability to anticipate the time that any given item would appear. Ts could be collected while they were occluded. However, they could only be found when they were visible. Thus, we could use the history of stimulus visibility for each clicked item to infer the time that it was found, relative to the time of the corresponding click. For example, if an item became invisible 500 msec before it was successfully collected, it *must* have been identified at least 500 msec before that collection (unless it was the unlikely subject of a lucky guess). For each click, we aligned the time course of visibility to the time of the click on the current item (Figure 6A). This could be done for either the current or next clicked item (upper and lower panels in Figure 6A, respectively). By summing across many target clicks, we can produce a distribution, showing the probability of visibility as a function of time before the click on the Current item (see Figure 6B and Methods, below). The peak of the resulting visibility distribution relative to the time of the click indicates when items were most likely to be visible. This can be taken as an estimate of when the items were found. This method has an advantage over color reports in that we can measure finding times throughout the trial efficiently, rather than just at the end of a trial. Moreover, no secondary task is required. Unlike eye tracking measures, this method does not rely on overt fixation on the T. This method would be sensitive to an observer's plans to click on a T, visible and attended to in the periphery.



**Figure 5. Procedure for Experiment 3.** (a) On each trial, observers continuously collected ‘T’s among distracting L’s. Each letter was intermittently occluded by gray squares. Each trial ended once either all Ts had been collected, or observers could choose to advance to the next trial once they had collected at least two Ts. (b) All items were occluded with a regular on-off cycle (e.g., 1000 ms visible, 3000 ms occluded), each with an independent, random phase. (c) Two additional visibility conditions were: 4000 ms visible / 4000 ms occluded, and 100% visible (no occlusion).

Experiment 3 consisted of two sub-experiments. In Experiment 3a, observers completed a foraging task similar to Experiments 1 and 2, under three visibility conditions: 1s-on/3s-off, 4s-on/4s-off, and continuously visible (Figures 5B and 5C). In addition, we varied target density, to determine how finding times varied as a function of the proportion of targets remaining in the display. Based on our results from the previous experiments, we predicted that on average, the peak of the visibility distribution of the next clicked target would fall before the time of the current click (0 ms). Experiment 3b served as a control experiment, in which the stimuli and procedure were identical, with the exception that on each click, all items were randomly shuffled among the remaining positions. Therefore, any information about future targets would become useless immediately after the click on the current target. Since observers would be forced to re-start the search after each click, the visibility distribution from this control provides an important point of comparison for Experiment 3a. It represents the pattern of results when observers cannot plan ahead. In Experiment 3b, we predicted that the peak of the visibility distribution of the next clicked target would be well after the time of the current click (0 ms).

## Observers

Experiments 3a and 3b had twelve participants each (respectively, 8 and 7 female; mean ages 29.5 and 32.4, ranges 18-54 and 20-55).

## Stimuli

The experiments were run on a testing configuration identical to Experiment 1. Search displays consisted of a total of 24 items, arranged into 4 rows  $\times$  6 columns, with 4, 7, or 10 ‘T’ shapes among ‘L’ shapes. The array was distributed across an area  $14.1^\circ$  wide  $\times$   $12.8^\circ$  high, prior to adding a random spatial jitter of up to  $\pm 1.02^\circ$  horizontally and  $\pm 1.73^\circ$  vertically on each trial. Items were drawn on a black background and were  $0.8^\circ$  in height and width ( $0.24^\circ$  line width). Items were shown in white for the full duration of the trial, with no change in color. Instead, all items within the display (both Ts and Ls) were intermittently occluded, with a regular on-off cycle, by gray squares ( $0.8^\circ$ ), and each item had a randomly assigned phase. Occlusion frequency and duration were manipulated to produce three stimulus Visibility Conditions: (a) 1 second visible / 3 seconds occluded (25% duty cycle) (b) 4 seconds visible /

4 seconds occluded (50% duty cycle), and (c) items continuously visible (100% duty cycle; no occluders).

## Procedure

Observers were instructed to click on the Ts as quickly as possible, regardless of whether they were visible or occluded at the time of the click. They were told that it would be necessary to click on occluded items in order to maintain a fast collection rate. Trials terminated in one of two ways: observers could click on all the T shapes, or they could choose to proceed to the next display once they had collected a minimum of two Ts. On each trial, once two Ts had been clicked on, a gray frame was added (filling the display except for a central 20.5° square region), and observers could click anywhere in the gray region to proceed to the next display. Observers were instructed to use whichever strategy allowed them to collect the Ts at the fastest rate. In both cases, a new display appeared automatically following a 500 ms blank intertrial interval (ITI). The number of Ts (4, 7, or 10) was randomly selected on each display.

Observers collected 600 Ts in each of the three Stimulus Visibility conditions, for a total of 1800 Ts in the experiment. The three conditions were completed in separate blocks, in a counterbalanced order across observers. Observers were given breaks after every 40 Ts collected. During the break, they were shown a display listing the total number of Ts collected, their collection rate since the last break, and their overall collection rate for the experiment. Observers then clicked the mouse button when ready to continue. At the beginning of the experiment, observers completed a short practice block in which they collected 40 Ts in each of the three conditions.

Experiments 3a and 3b were identical except for a shuffle manipulation that was added to Experiment 3b. In Experiment 3b, on each click (both hits and false alarms), all the remaining Ts and Ls on the display (both occluded and unoccluded) were randomly shuffled among the remaining item locations, with the constraint that no location could be a T before and after the click. Occluders moved together with the items, so that the shuffle manipulation did not change the visibility of the item (i.e., an item that was occluded before the shuffle was still occluded after the shuffle, at its new location).

In both experiments, auditory and visual feedback was provided to encourage fast collection while discouraging false alarms. When a 'T' was clicked (either occluded or unoccluded), a rising tone was played to indicate a correct response (250 to 1000 Hz quadratic frequency sweep; 250 ms duration). The 'T' disappeared following a brief (67 ms) animation of an expanding green ring (1.13° to 2.09° diameter) over the item. When an 'L' was clicked, a buzzing sound was played (generated from the absolute value of the sum of 140, 280, and 560 Hz pure tones; 250 ms). A red 'X' (1.7°) was also briefly

drawn over the item (in its occluded or unoccluded state) for 67 ms, but the item was not removed from the screen.

To further incentivize rapid collection, observers were given a payment bonus based on their overall collection rate. In Experiment 3a, observers were given an additional \$1 per every increase of 0.1 items/second above 1.25 items/second (inclusive), added to \$11 for one hour of participation (with a cap of \$14). In Experiment 3b, in anticipation of slower collection rates due to the shuffle manipulation, observers were given an additional \$1 per every increase of 0.1 items/second over 0.75 items/second. Collection rates in Experiments 3a and 3b were  $1.20 \pm 0.16$  and  $0.81 \pm 0.13$  items/second ( $M \pm SD$ ), respectively.

## Data Analysis

Visibility distributions were calculated separately for Current and Next targets by aligning the timeline of visibility (for the currently clicked item or the next clicked item, respectively) to the time of the click on the current item. From these timelines, we calculated the proportion of targets, across all clicks, that were visible at each frame, from -3000 to +3000 ms relative to the time of the current click (0 ms). Note that these distributions reflect the fact that targets disappear as they are clicked. Thus, the visibility of Current targets drops to 0 at the time of the current click at 0 ms because the item vanishes at that point. In addition, we only analyzed clicks on Ts (hits). Clicks on Ls (false alarms), which accounted for 3.0% and 5.1% of all clicks in Experiments 3a and 3b, respectively, were removed from the analysis. Distributions were calculated separately for the two occlusion conditions (4s-on/4s-off and 1s-on/3s-off) and could be separately analyzed by the number of targets remaining at the time of the click (1-10 targets for Current and 2-10 targets for Next analyses). The continuously visible condition only provided a point of reference for target collection rates, and was not analyzed otherwise. From each distribution, we calculated the time of the peak closest to the time of the click, as an estimate of the approximate time that items were found (see Discussion and Supplemental Note 3 for further details).

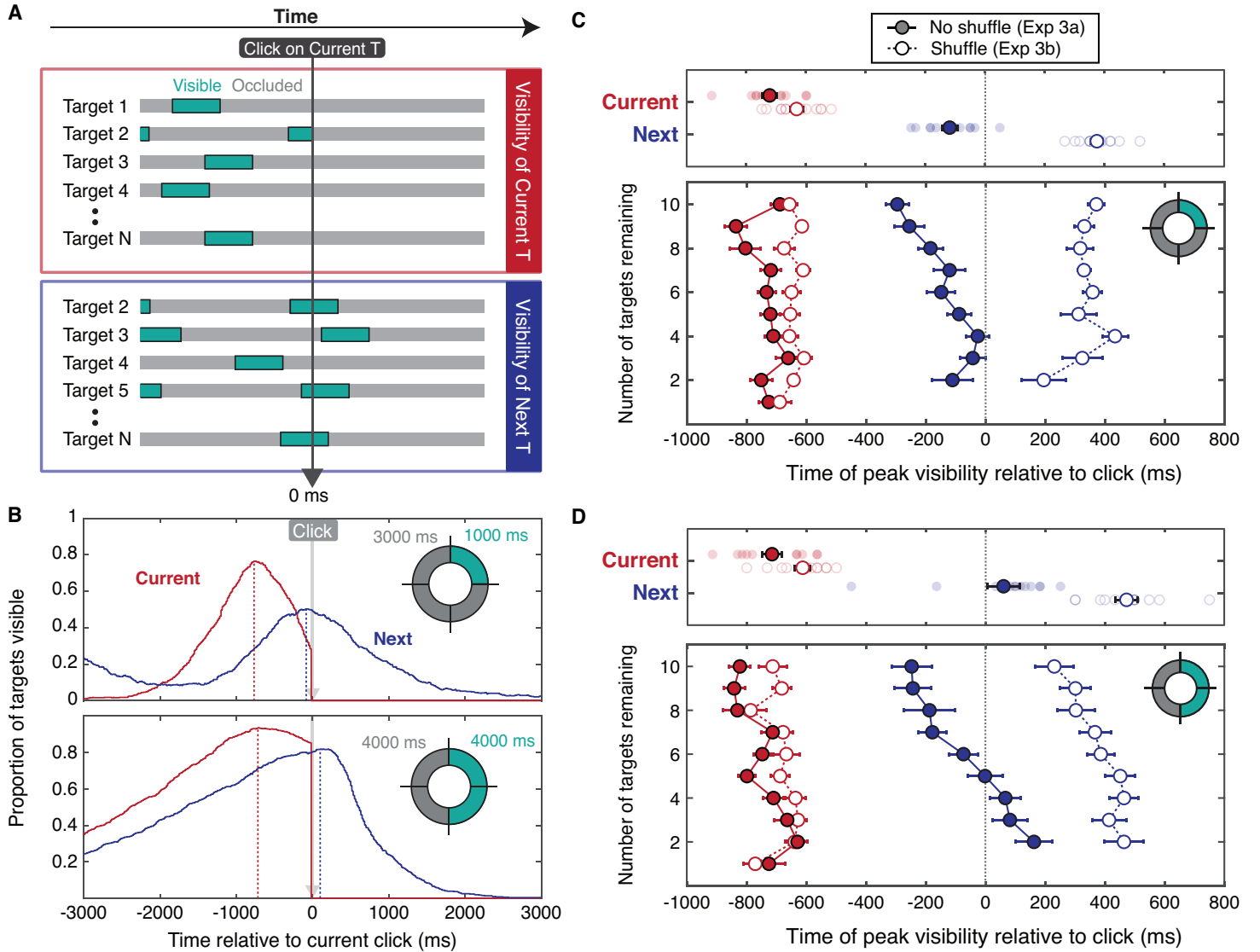
## Results

### Collection rates

Collection rates, calculated from mean inter-click intervals for each condition and each observer, were analyzed with a 2 (Experiment: Shuffle vs No-shuffle) x 3 (Visibility Condition) mixed-model ANOVA, with Experiment as a between-subjects factor, and Visibility Condition as a within-subjects factor. There was a significant main effect of Visibility Condition,  $F(2,44) = 65.89$ ,  $p < .001$ ,  $\eta_p^2 = .75$ . Pairwise post-hoc contrasts using Tukey's HSD tests showed significant differences in mean inter-click



interval between each pair of Visibility conditions (all  $p$ -values  $< .001$ ). Collection rates were fastest in the continuously visible condition, followed by the 4s-on/4s-off, and the 1s-on/3s-off condition (mean inter-click intervals were 912, 1060, and 1218 ms, respectively, averaged across the Shuffle and No Shuffle conditions). There was also a significant main effect of Experiment, with slower inter-click intervals when items were shuffled compared to when they were not (1275 vs 851 ms, averaged across visibility conditions),  $F(1,22) = 33.7$ ,  $p < .001$ ,  $\eta_p^2 = .61$ . The Experiment  $\times$  Visibility Condition interaction was not significant,  $F(2,44) = 1.26$ ,  $p = .29$ ,  $\eta_p^2 = .05$ .



**Figure 6.** (a) For each click, we aligned the pattern of visibility to the time of the click on the respective target (Current analysis, upper panel), or the pattern of visibility of the next clicked item to the click on the current target (Next analysis, lower panel). Note that targets visible at the time of the click disappear following the click (e.g., target 2 in upper panel). (b) Visibility distributions for one representative observer, for the 1s-on/3s-off condition (upper panel), and the 4s-on/4s-off condition (lower panel). These were calculated by averaging, across clicks, the pattern of visibility of the item relative to the time of the click. Red and blue curves show the proportion of Current or Next clicked targets that were

visible around the time of the current click. Dotted vertical lines show the peak of each distribution, (c) Time of peak visibility in the 1s-on/3s-off condition, calculated from distributions collapsed across all clicks (upper panel), and calculated from separate distributions for each number of targets remaining in the display (lower panel). (d) Same as (c), for the 4s-on/4s-off condition. Error bars in (c) and (d) represent  $\pm 1$  SEM.

## Visibility distributions

Figure 6B shows sample visibility distributions for a representative observer, for both Current and Next targets, in both the 1s-on/3s-off and the 4s-on/4s-off conditions. For each observer and condition, we calculated the peak of the visibility distribution as an estimate of target finding time, relative to the time of the current click (Figures 6C and 6D). Distribution peaks were compared between conditions with a 2 (Experiment: Shuffle vs No-shuffle)  $\times$  2 (Visibility Condition)  $\times$  2 (Target Type: Current or Next)  $\times$  9 (Targets Remaining) mixed-model ANOVA, with Experiment as a between-subjects factor and the remaining factors as within-subjects factors. Where reported, post-hoc comparisons were performed using Tukey's HSD tests. The results are summarized briefly here and reported in full in Supplemental Note 3.

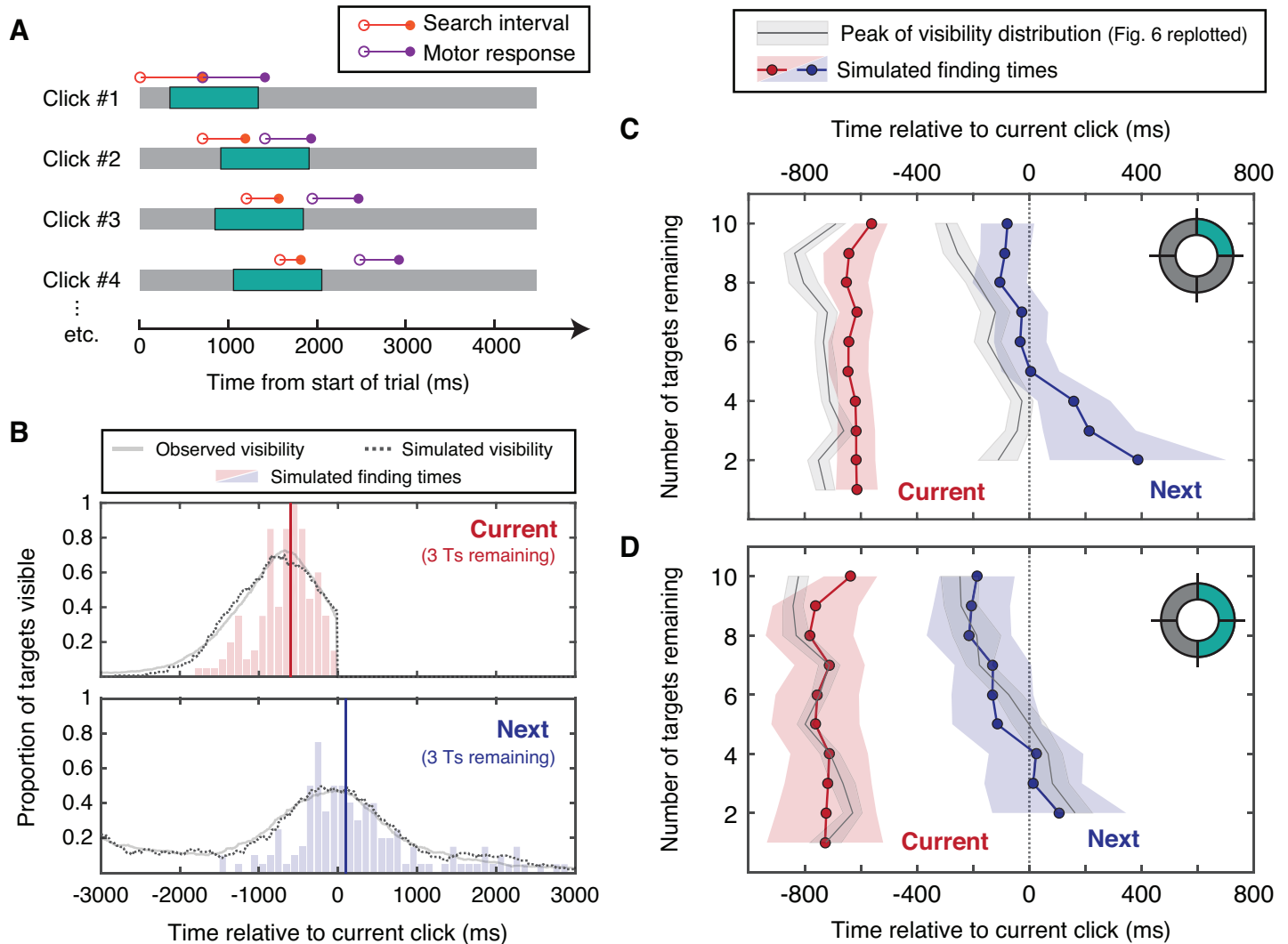
As shown in Figures 6C and 6D, the peak of the visibility distribution for the Current item falls approximately 700 ms before the current item is clicked, consistent with the results of the previous experiment. In addition, the peak of the Next distribution falls roughly at the time of the click on the Current item. This difference is reflected in a significant main effect of Target Type (Current vs. Next);  $F(1,22) = 1142.29, p < .001, \eta_p^2 = .98$ . There was also significant interaction between Experiment and Target Type,  $F(1,22) = 58.87, p < .001, \eta_p^2 = .73$ . As predicted, the shuffle manipulation—in which observers were required to re-start the search after each click—shifted the peak of the distributions for Next targets much later (a mean value of 353 ms, compared to -105 ms without the shuffle;  $p < .001$ ), and shifted the Current distribution only slightly later, from -743 ms to -660 ms ( $p = .01$ ). Finally, as shown in Figure 6, there is also a clear effect of the number of Ts remaining on the peak of the Next distribution, reflecting the greater difficulty in finding a rarer target. Consistent with this observation, both the main effect of Targets Remaining ( $F(8,176) = 16.42, p < .001, \eta_p^2 = .43$ ) and the interaction between Targets Remaining and Target Type ( $F(8,176) = 4.00, p < .001, \eta_p^2 = .15$ ) were significant. In other words, as the number of Targets Remaining decreases, observers find the Next item at progressively later time points, and in some cases after the click.

## Discussion

In Experiment 3, we developed an intermittent occlusion paradigm in which we systematically varied item visibility, inferring the times that targets were typically found based on the times that they were most often visible. Our results can be summarized as four main findings. First, on average, observers *found* the Next clicked item slightly before, or close to, the same time as they *clicked* on the Current item. In Experiment 3a, averaged across Visibility Conditions, the peak visibility of the Next clicked item was -105 ms relative to the click on the Current item, consistent with the pattern of results in Experiments 1 and 2 (-175 ms to -157 ms for color reports, and +48 ms for fixation onsets across all clicks). Second, within each trial, as the number of targets remaining decreased, it took longer for observers to find the next target, reducing observers' ability to find additional Ts ahead of each click. This is consistent with the observed shift in the peaks of the visibility distributions (Figure 6C and 6D; bottom panels) toward time points after the time of the click. We note that variation in the initial number of targets (4, 7, or 10) may also contribute to this shift toward progressively later target finding times (see Figure S6). Third, our results are generally robust over differences in the rate of occlusion; as shown in Figures 6C and 6D, peak finding times are similar between the 1s-on/3s-off and the 4s-on/4s-off cycles, suggesting that these results are unlikely to be an artifact of the specific occlusion rate selected. Fourth, we find that in Experiment 3b, when we shuffle items between the remaining locations on each click, observers are forced to re-start the search each time. Therefore, the peaks of the visibility distributions for Next targets are shifted much later, to approximately 400 ms after the click. This further verifies that the results we observe in Experiment 3a are not due to an artifact of the experiment design or analysis procedure, and instead reflect the time that targets are found. Together, these results support the validity of this new procedure, and replicate the target finding times observed in the other two experiments.

While this new technique offers some advantages over color self-reports and eye tracking, it relies on the assumption that the peak of the visibility distribution is closely related to finding time. Therefore, we separately evaluated the degree to which the time of peak visibility reflects the time that targets were found using simulations of the task (see Supplemental Note 4 and Figure 7 for details). As shown in Figure 7A, we simulated finding times using a procedure in which unoccluded items were sequentially 'checked' until a target was found, with a delay added for each checked item. A motor response delay was then added to the finding time to estimate the time when each target would be clicked. To simulate observers' ability to plan ahead, additional targets could be searched for and found during the motor response intervals for previous targets. By systematically varying the simulation parameters (in over 35,000 unique combinations), we produced visibility distributions that closely matched those in the observed data set in Experiment 3a. Next, we calculated the median of the

underlying finding times from each simulation (Figure 7B). Figures 7C and 7D show the finding times from the average of all the best-fitting simulations to the observed data (1<sup>st</sup> percentile of sum-of-squared errors), along with the original data, replotted from Figure 6. Here we see that the peaks of the visibility distributions closely approximate the underlying finding times. Peak visibility slightly underestimated finding times in the 1s-on/3s-off condition (difference of -317 ms to -85 ms, averaged across Current/Next) compared to the 4s-on/4s-off condition, which was approximately centered on the observed data (-123 ms to +75 ms). Nevertheless, the results of the simulations support our main conclusions from Experiment 3—that observers typically find the Next item close to the time of the click on the Current item, and that these finding times occur later as targets become sparser.



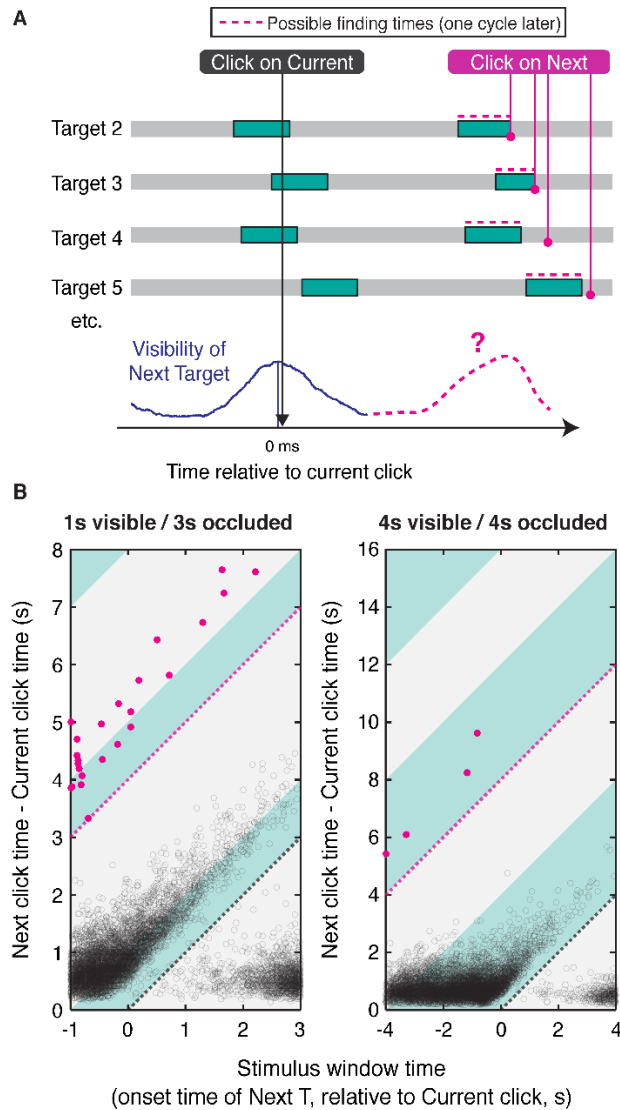
**Figure 7.** Simulations comparing peak visibility to finding time in Experiment 3A. (A) Timelines of visibility (as in Fig. 6A) aligned to the start of the trial for a set of four targets. Targets are sorted by the order in which they were clicked, from top to bottom (Click 1, 2, etc.). Orange lines above each target represent the time spent searching for that item (including the time spent checking distractors), and purple lines represent the time spent collecting the item (motor response). Open and filled circles represent the start and end of each interval, respectively. Filled orange circles represent finding time and

filled purple circles represent the time of the click. Note that items can only be found when they are visible. In addition, search intervals cannot overlap with each other, and motor response intervals cannot overlap with each other; however, search intervals may overlap with motor responses for previously found targets. (B) Comparison between visibility distributions and finding times. Solid light gray line shows observed visibility distribution, calculated from the average of all observers' visibility distributions (e.g., Figure 6B) within the same condition (e.g., 3 Ts remaining). These observed visibility distributions were then compared to the visibility distributions produced by the simulations (dotted gray line). In this instance, the simulated visibility distribution closely matches the observed data. The corresponding histograms show the underlying finding times from the simulation (i.e., filled orange circles in panel A). Current and Next analyses are shown in red and blue, respectively, and solid vertical lines represent median finding times. (C, D). Comparison between simulated finding time and observed data (peak visibility), for the 1s-on/3s-off condition (C) and the 4s-on/4s-off conditions (D). Gray shaded areas represent observed data in Experiment 3a (replotted from Figures 6C and 6D, where error bars represent  $\pm 1$  SEM). Red and blue lines represent the average finding times from the best-fitting simulations to the observed data (1<sup>st</sup> percentile of sum-of-squared errors), with shaded regions representing  $\pm 1$  SD.

A further consideration in using cyclical on-off occlusion to infer target finding times is that items are visible at regular, repeated intervals throughout the trial. To illustrate this, Figure 8A shows the visibility of the Next clicked item, aligned to the click on the Current item. This is similar to the lower panel of Figure 6A, but extended forward in time. In this diagram, Next items tend to be visible close to the time of the Current click, and in interpreting the results of Experiment 3, we assume that items are found during this interval. However, as shown in Figure 8A, items that have not been clicked yet will re-appear several seconds later. In these cases, the Next item will be visible multiple times in between the Current and Next click, and observers could be finding Next clicked items either: (1) close to the Current click, or (2) one cycle *later* (i.e., closer to the time of the Next click). In the latter scenario, observers might be waiting for targets to re-appear after they are occluded, and this would be inconsistent with our conclusion that observers are planning ahead. To exclude this possibility, we calculated the proportion of targets that re-appear before they are clicked. Figure 8B shows the times that targets are clicked as a function of the stimulus window. It shows that nearly all clicks occur before the target is visible for a second time. Summed across the 1s-on/3s-off, and the 4s-on/4s-off conditions, the cases in which the target is visible for a second time following the click account for 0.19% of clicks in Experiment 3a (Figure 8B) and approximately 1.03% of all clicks in Experiment 3b (Figure S4). This indicates that observers rarely wait for targets to re-appear, and that targets are typically found during the cycle closest to the Current click.

Finally, we note that the opposite situation (from that shown in Figure 8A) could be true –Next items could be found one cycle *earlier*. However, in this case, the result would be consistent with observers planning ahead, and this would mean that our estimate of finding time for Next items (-105 ms

relative to the click on the Current item) is somewhat conservative. Nevertheless, we calculated the total proportion of all clicks in which the target is visible for more than one interval during the trial (see Figure S5). This is similar to the analysis in Figure 8B, with the exception that click time and stimulus window time are referenced to the start of the trial (rather than the preceding click). This represents a somewhat larger proportion of clicks in Experiment 3a (24.3% and 10.4% of all clicks in the 1s-on/3s-off and the 4s-on/4s-off conditions, respectively). We re-calculated the visibility distributions after removing these clicks from the analysis, and found a similar pattern of results (Figure S5). Together, these additional analyses indicate that items are most likely found during the cycle closest to the time of the click.



**Figure 8.** Analysis showing target visibility in-between clicks for Experiment 3a. (A) Illustration of a hypothetical scenario in which Next items are visible near the time of the Current click, but are found several seconds later, when they are visible again. Turquoise and gray regions represent periods of visibility and occlusion, respectively, aligned to the time of the Current click. In each case, the item

could be found during the re-appearance of the target (magenta dashed line) before it is clicked (magenta circles). (B) Scatterplot of click time with respect to item appearance for all targets collected, across all observers in Experiment 3a, shown separately for the 1s-on/3s-off and the 4s-on/4s-off conditions (left and right panels). The y-axis shows the time interval between the Current and Next click. The x-axis shows the time point within the stimulus visibility cycle corresponding to the Next clicked item (i.e., the onset time of the Next target, relative to the Current click). Negative values (-1 to 0) represent Next targets that were already visible at the time of the current click (e.g., targets 2, 3, and 4 in panel A). Positive values represent Next targets that were occluded during the click on the current item (e.g., target 5 in panel B). The black open circles represent clicks that occurred before the Next target re-appeared during the trial (i.e., the target was visible no more than once). These also include instances in which the target was fully occluded in-between clicks, consistent with the target having been found ahead of the current click (data points in the lower-right corner, below the black dotted line). Data points above the magenta dotted line (filled magenta circles) are clicks that occurred *after* the target re-appeared during the trial (examples shown in panel A; 0.19% of clicks in Experiment 3a).

#### **Experiment 4: How far ahead can observers plan?**

Experiments 1-3 provide converging evidence to indicate that observers can plan ahead while foraging. Next, we tested how far ahead observers can plan beyond the current item. In Experiment 1 and Experiment 2a, observers only indicated the location of the *one* item to be clicked next. However, introspectively, it can feel like we know about the presence of several targets. If we occlude the items at an unpredictable time point during the search, and ask observers to click on all the targets that they know about, it seems likely that they would be able to correctly identify more than one target on a sizable proportion of the trials. How many targets can observers get? The number of targets observers can correctly identify is likely to depend on how long they are able to view the display. For example, an observer is not likely to get very many targets correct if they view the display very briefly (e.g., 100 ms). However, if we allow them to forage for a few seconds, they may get multiple targets. In other words, observers may need to forage for some time before they reach an asymptotic level of performance. What does this time course look like? In Experiment 4, we measured the limits of observers' ability to search ahead, by randomly interrupting their search for Ts, and asking them to click on the locations of the remaining Ts.

As shown in Figure 9, we measured observers' capacity for collecting future targets in two conditions. In one condition (referred to as "No Shuffle" trials), observers continuously foraged for targets as before, and all items were abruptly and unpredictably occluded 100 – 2500 ms following display onset. Secondary to the aim of this experiment, we also included a "Shuffle" condition, in which all items were randomly shuffled between the remaining locations on the 3<sup>rd</sup>, 4<sup>th</sup>, or 6<sup>th</sup> click, forcing observers to re-start the search. At a random interval 100 – 2500 ms later, all remaining items were occluded. In both conditions, at the onset of the occluders, observers were instructed to indicate the

locations of targets, with the ability to select multiple boxes. As we mention above, the No Shuffle condition provides a measurement of the buildup of future target planning at the onset of the trial. The Shuffle condition is similar, but measures the time course of collection of future targets, after the search is disrupted. Importantly, unlike Experiment 3, the shuffle manipulation in this experiment was not intended to distinguish planning from no planning (as the previous experiments already established this). Presumably, if we shuffle all the items in the display and then terminate the trial immediately, the number of hits will be near chance (i.e., similar to viewing the display for only 100 ms). If we shuffle all the items in the display, and then allow observers to forage for a few seconds, we would expect them to get one or more targets. How long after the disruption does it take for observers to reach the same level of performance again? In each condition, we compared performance to the number of Ts expected by random guesses. Based on the results of the previous experiments, we expected that observers would be able to accurately identify at least one future target, and would be above chance in identifying Ts beyond that single item.

## **Method**

### **Observers**

Twelve observers (4 female; mean age: 30.6, range: 20 – 53) participated in Experiment 4.

### **Stimuli and Procedure**

The testing configuration and stimulus parameters were similar to Experiment 3, with the following exceptions. Items were fully visible for the duration of the trial until it was unpredictably interrupted. The time of the interruption was determined by one of two types of trials: No-Shuffle trials and Shuffle trials. On No-Shuffle trials, the display stopped at one of 8 random time points relative to the start of the trial, logarithmically spaced between 100 and 2500 ms. On Shuffle trials, all the Ts and Ls were randomly switched among the remaining locations at the onset of the 3<sup>rd</sup>, 4<sup>th</sup>, or 6<sup>th</sup> click. The trial then terminated between 100 and 2500 ms later. The two conditions were randomly interleaved, and observers were not told which trial type they were completing. Mixing Shuffle and No Shuffle trials should discourage observers from pausing at the outset of a trial to try to memorize multiple locations. The starting number of Ts was varied to approximately equate the target density at the end of the trial between the Shuffle and No Shuffle conditions. The Shuffle condition had 3 densities: 9 Ts initially (exit after 6<sup>th</sup>), 9 Ts (exit after 4<sup>th</sup>), and 10 (exit after 3<sup>rd</sup>). The No-Shuffle condition had 3 densities: 4, 6, or 8 Ts initially. Expressed as a proportion of the targets remaining at the moment of occlusion, the resulting target densities were the same for the Shuffle and No-Shuffle conditions: 16.7%, 25% and 33.3%.



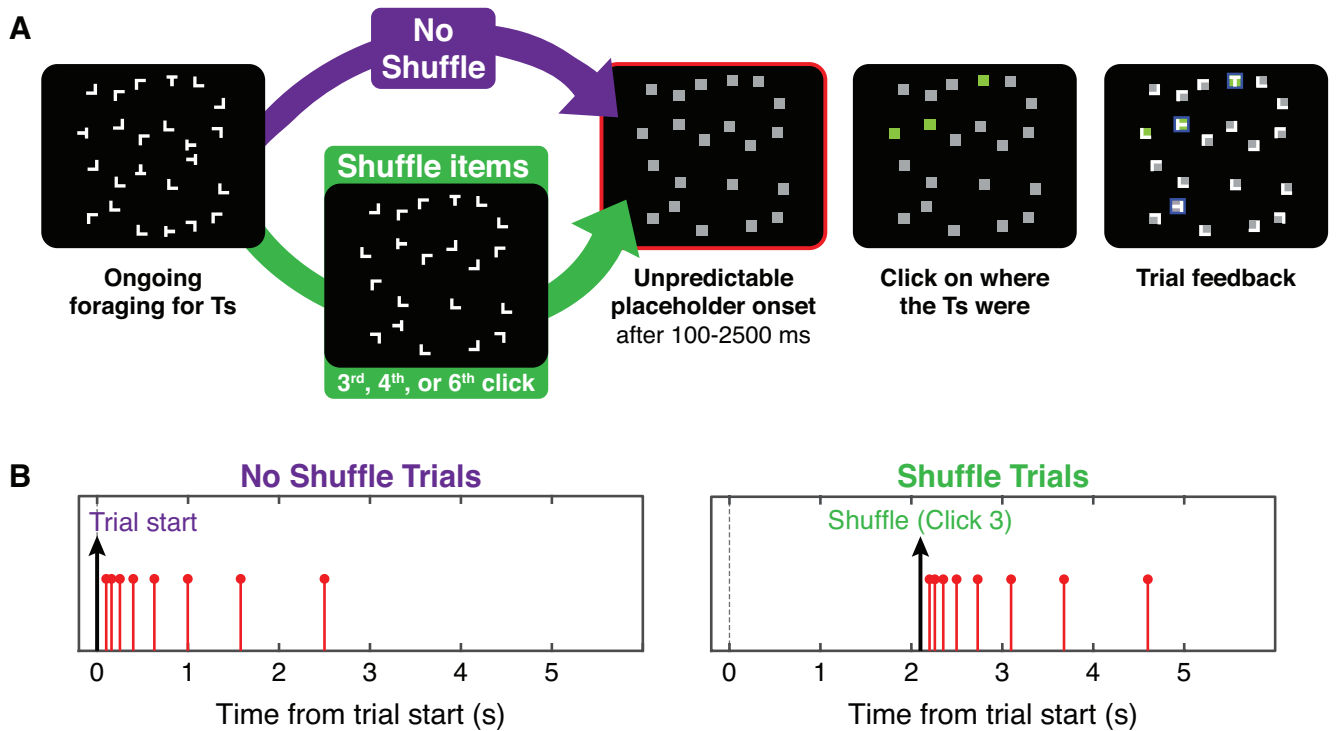
At the end of each trial, the remaining items were replaced by gray squares ( $0.8^\circ$ ). Observers were instructed to click on the boxes corresponding to Ts that they were planning on clicking next. Boxes that observers had clicked on changed to green, and observers could de-select boxes with a second click. Observers were told that they could click on multiple boxes, or not click on any boxes. Once finished with their selections, observers clicked a button labeled “Done” at the bottom of the screen, and were then shown a feedback screen showing the identities of all the items under the occluders, which boxes contained Ts (highlighted with a blue frame), the number of correctly identified Ts, and the number of false positives. Observers then clicked the mouse button when ready to continue to the next trial.

To make the shuffle less noticeable, a brief “twitch” was added that coincided with each click, in both the Shuffle and No-Shuffle trials. At the onset of the click, all items briefly shifted in location, each by a random spatial jitter of up to  $\pm 0.28^\circ$  horizontally and vertically, and then shifted back after 50 ms. As before, the T-shapes disappeared as they were clicked, and observers were given auditory feedback similar to Experiment 1 to encourage rapid collection. Observers heard a tone at regular intervals if the time from the previous click on a T shape (or from trial onset) exceeded 1500 ms. No other auditory or other visual feedback was provided.

Observers completed 10 trials for each unique combination of Trial Type (Shuffle or No-Shuffle), Inter-stimulus Interval (ISI; Stimulus/Post-Shuffle Duration, with 8 conditions) and Target Density (3 conditions), for a total of 480 trials, which were randomly interleaved. Observers also completed a short 20-trial practice block at the beginning of the experiment.

### **Data Analysis**

For each trial, we calculated the number of hits as the mean number of Ts selected on the placeholder screen, separately for each observer, ISI, and Trial Type. For each trial, we also calculated the number of hits expected by chance by multiplying the number of selected items by the proportion of targets out of the total number of items remaining in the display. As with the number of hits, the chance level of performance was calculated by averaging the chance number of hits for each observer, ISI, and Trial Type.

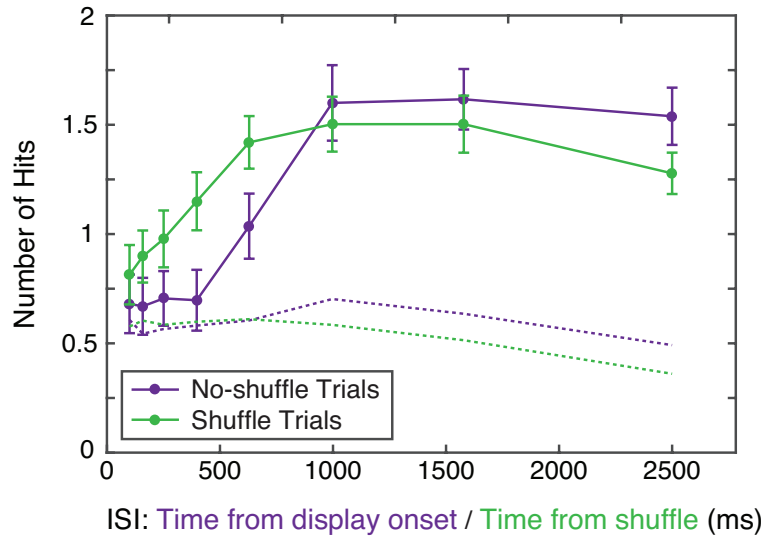


**Figure 9. Procedure for Experiment 4.** (a) Observers continuously foraged for Ts, as in Experiments 1-3. Trials were randomly assigned to one of two conditions, and observers were not told which condition they were completing in advance. In the No-Shuffle condition, the trial abruptly terminated after 100-2500 ms. In Shuffle trials, the trial abruptly terminated 100-2500 ms after the 3<sup>rd</sup>, 4<sup>th</sup>, or 6<sup>th</sup> click. At the end of the trial, all remaining items in the display were occluded by gray boxes, and observers were instructed to select where the remaining Ts were, with the ability to select multiple boxes. (b) Sample trial timelines for No Shuffle trials (left) and Shuffle trials (right), with 0 ms indicating the onset of the search display. Black arrows indicate either the time of the trial start (No Shuffle trials) or one of the clicks (Click 3; Shuffle trials). Red markers indicate possible occluder onset times.

## Results

Figure 10 shows the mean number of hits across all observers, as well as the group chance performance level, at each ISI, separately for the No-Shuffle and Shuffle conditions. A 2 (Trial Type) x 8 (ISI) repeated-measures ANOVA on the number of hits showed a significant main effect of ISI,  $F(7,77) = 67.36$ ,  $p < .001$ ,  $\eta_p^2 = .86$ . Averaged across Trial Type, the mean number of hits ranged from 0.74 to 1.56. There was also a significant main effect of Trial Type,  $F(1,11) = 21.79$ ,  $p < .001$ ,  $\eta_p^2 = .66$ , with a greater overall number of hits in the Shuffle condition compared to the No-Shuffle condition (1.19 vs 1.06, respectively). This difference was driven by an increased number of hits in Shuffle trials at short ISIs (less than 1000 ms), as indicated by a significant Trial Type x ISI interaction,  $F(7,77) = 16.18$ ,  $p < .001$ ,  $\eta_p^2 = .60$ . In addition, at each ISI, we performed pairwise comparisons between the mean number of hits and the expected chance performance level, which was calculated separately for

each observer. In the No-Shuffle trials, performance was significantly above chance only at the longest four ISIs,  $t(11) > 5.77$ ,  $p < .001$ , at a Bonferroni-corrected alpha ( $\alpha_B$ ) of .003. In the Shuffle trials, performance was above chance at all eight ISIs,  $t(11) > 4.39$ ,  $p \leq .001$  ( $\alpha_B = .003$ ).



**Figure 10. Results from Experiment 4.** Mean number of hits (correct T selections) at each interstimulus interval (ISI), shown separately for Shuffle and No-Shuffle trials in green and purple, respectively. For Shuffle trials, the x-axis represents the shuffle-to-placeholder ISI. For No-Shuffle trials, the x-axis is time from stimulus onset to placeholders (in the timelines in Figure 9B, the x-axis corresponds to separation between the black arrow and red markers). The dashed lines represent the group chance performance level for the two trial types, averaged across all observers. Error bars represent  $\pm 1$  SEM.

## Discussion

In Experiment 4, we measured observers' capacity for collecting future targets by testing whether they could accurately indicate the locations of multiple Ts after the trial was abruptly terminated. Results from both the Shuffle and No-Shuffle conditions demonstrated that observers accurately identified the locations of approximately 1.5 future targets at their asymptotic level of performance, planning target collection 1-2 items ahead. In the No Shuffle condition, the display needed to be shown for at least 1000 ms for observers to reach that level of performance, and needed to be shown for at least 630 ms for observers to perform above chance. In the Shuffle condition, items were randomly switched between the remaining locations, and we measured how long it took observers to re-acquire the same number of targets. Compared to the No Shuffle condition, observers needed slightly less time to reach the asymptotic level of performance after the search was interrupted, acquiring a similar number of targets in the last 630 ms before the placeholders appeared. In addition, observers

were above chance in identifying Ts with very short ISIs, performing above chance when the display was occluded 100 ms after the shuffle.

What accounts for this shift in performance between the Shuffle and No Shuffle trials? In the Shuffle trials, items are switched between the same set of locations in the display, rather than being shown in a new spatial configuration. Therefore, one possibility is that there may be some benefit to observers being already familiar with the spatial layout of the display. Another possibility is that there may be a general benefit to observers already being in the middle of the task when items are shuffled. In other words, observers might be generally slower at the outset when shown a new display, taking longer to acquire new items. This is also broadly consistent with the observation that the interval from display onset to the first click is typically longer than the intervals between subsequent clicks. Regardless, it seems that there is a smaller penalty in collecting targets, if observers are interrupted mid-search, compared to when they are given a new display to start from. More generally, our findings here are consistent with results in the previous three experiments, demonstrating that observers can typically forage 1-2 items ahead.

### **General Discussion**

Most of the literature on visual search (indeed, most of the experimental literature in perception and cognition) describes experiments with a discrete trial structure. This makes sense since we want multiple measures of the same task in order to extract a signal from the noise of trial-to-trial, moment-to-moment variability. However, the processes that are studied in this way in the lab are typically embedded in a continuous stream of behavior in the real world. Important properties of that embeddedness can get lost when the task is examined in discrete trials. In search, trial structure forces one search to end before the next one begins. That limitation can be overcome by using multiple target, foraging tasks. Using those tasks allows us to document the main message of our current work – that the target of the next search can be found before the current search is completed.

Using a combination of behavioral and eye tracking methods, including new procedures that we have introduced (e.g., color report and intermittent occlusion), Experiments 1-3 measured the time when targets were found within a continuous foraging task. Those times to find the next target were typically before (or around the same time) as the click on the current item, indicating that observers generally plan ahead when collecting multiple targets and that any interference between target collection and the search for new items is likely to be minimal. In addition, Experiment 3 shows that the ability to find future targets ahead of the current click becomes more limited as items become sparser in the display. Although these estimates vary with the difficulty of the search task, the different methods used here

converge on a consistent, roughly 200-ms range for the typical finding time for the next item, relative to the current click (0 ms): -175 ms for color reports, -105 ms for intermittent occlusion, and +48 ms for fixation onsets. Some of these values are likely constrained by the specific methods used. For example, the color reports are likely at the lower bound of this range, because the values are limited to times before the end of the trial (i.e., before the last click), and therefore must be negative. Fixation onset time, on the other hand, might produce different and later estimates, because gaze behavior will not register information about items that are covertly attended to and found at locations away from the point of fixation. Importantly, despite these variations in target finding time, these finding time estimates are very different from those that would be expected if observers were waiting to complete collection of the current item before starting the search for the next item. Experiment 3b provides a point of comparison, because items were shuffled on each click, forcing observers to re-start the search each time. Here, finding times are much later—on average, 353 ms after the time of the click.

We note that the values reported here likely represent a conservative estimate of observers' ability to search ahead for future targets for a couple reasons. First, these experiments estimate the time when the next target is *found*. The next search must have *started* some time before that. Whether a search for a T among Ls involves serial deployments of attention from one item to the next (e.g., Kwak, Dagenbach, & Egeth, 1991) or some more parallel process (e.g., Sung, 2008), the search could have begun several hundred milliseconds before the time that the T was found. This is particularly relevant for the late estimates of finding time that we observed when there were few items remaining in the display (Experiment 3). It is unlikely that observers switch strategies and stop searching for future targets entirely; a more plausible explanation is that observers continue searching ahead, and find targets later simply because they are scarce. Moreover, the methods described here probably provide estimates of the most recent time that a target was found or attended. It is possible that the T was found, remembered, and then re-attended when it was time to commit to the collection of the Next target. Thus, again, the estimates from our experiments of when observers search for and find the Next T are probably quite conservative. It seems reasonable to assume that much of the time before the collection of the current target is spent searching for and planning collection of the next target.

In Experiment 4, we also report that there is a limit to the number of items that observers plan for future collection. We show that, given enough time to view the display (1000 ms or more), observers can typically plan collection 1-2 items ahead, and this result generally agrees with previous work. Thornton and Horowitz (2004) showed that when uncollected items in a display are shuffled, observers are considerably slower in collecting future targets, as measured by the interval between successive clicks. However, when the next target is left in place (and the remaining items shuffled) observers'

collection rates are much closer to typical (unshuffled) rates. This advantage seems to primarily occur for the next target; keeping the targets two or more steps ahead in place confers a smaller advantage in collection time.

While this is consistent with our finding that observers search ahead for future items, these capacity estimates are lower than typical estimates of visual working memory capacity, when expressed as a number of discrete items or chunks (e.g., 3-4 items). If there is minimal interference between the search for new items and target collection, why do we not observe a larger capacity? First, we note that these working memory limits are not absolute, and that information can be represented in working memory in different ways (see Brady et al., 2011 for a review). Moreover, the limit we measured likely reflects how quickly new targets can be found, relative to how quickly they can be collected. In Experiment 4, we see that the number of items that have been found (but not collected yet) levels off with longer durations. However, as trial duration increases, the density of targets present in the display also drops, and the search for new items takes longer. Therefore, had we tested higher target densities (relative to trial duration), we might have observed larger capacity estimates. Another limitation has to do with how we measured capacity. For a response to be considered a “hit”, observers needed to report the exact location of a T. However, it is possible that observers may know the approximate locations of future targets, beyond the 1-2 that we measured. Recent experiments on multiple identity tracking, for example, have shown that observers are able to report the approximate locations of tracked targets, well above typical capacity estimates obtained with all-or-none accuracy measures (Wu & Wolfe, 2018). Therefore, much like the estimates of finding time, our estimate of future target capacity is likely to be fairly conservative.

Even if these estimates are conservative, our results may account for the high level of efficiency reported in search studies with multiple targets. For example, previous work has shown that foraging search operates faster than what would be predicted from single-target searches, particularly during the ‘cruise’ phase in the middle of each trial (i.e., excluding the first and last targets; T. Kristjánsson et al., 2020). In agreement with this work, our results indicate that observers’ foraging strategies are distinct from those used when searching for a single target, and serve to maximize efficiency. A complete account of exactly how observers plan ahead during search would require further work, but it is likely that attention is distributed over multiple future targets as the observer collects each item. This is consistent with the literature on movement planning, which demonstrates that visual attention can be distributed over multiple future saccade targets (Baldauf & Deubel, 2008; Gersch, Kowler, Schnitzer, & Doshier, 2009; Godijn & Theeuwes, 2003), and that attention is deployed to multiple locations when planning reaching and grasping movements (see Baldauf & Deubel, 2010, for a review). Furthermore, in

visual search tasks, multiple saccades can be planned in parallel, and information is simultaneously processed at multiple future saccade endpoints (Caspi, Beutter, & Eckstein, 2004; McPeck, Skavenski, & Nakayama, 2000). While these mechanisms have been studied with single-target searches, it is possible that similar processes are involved here.

We know from these previous studies and from the dual task literature (e.g., Pashler, 1991) that we *can* do more than one thing at a time. Less is known about how observers choose to organize their search behavior in complex settings. Together, these experiments offer new approaches for examining continuous search behavior, and introduce new ways to probe finding time (e.g., occlusion, color reports) while observers search for multiple targets. Our foraging task is still a very limited approximation of the real world of continuous action, and we expect that these methods can be adapted to other types of experiments examining foraging behavior with more naturalistic tasks. For example, our foraging task is limited to a sequence of the same search task (finding T shapes). Would these results look similar if observers were performing a “hybrid search” task (Wolfe, 2012; Wolfe, Aizenman, Boettcher, & Cain, 2016), in which observers search for instance(s) of any of several different target types? For example, if you are currently collecting a nail, would you already be searching for a hammer (or something less related)? Similarly, our task is also limited to one specific type of motor response (clicking on targets). In the real world, appropriate motor responses vary considerably depending on the type of target (e.g., reaching to pick berries), and in many situations, the observer must also coordinate search with navigation through the environment (e.g., walking, driving). Future research examining how observers coordinate search with motor responses (target collection) and other tasks (e.g., navigation) would offer insight into the organization of continuous behavior within more naturalistic settings.

### **Pre-registration and data availability**

All data and materials are available on the Open Science Framework (OSF) online: <https://osf.io/zkyj2/>. Preregistration information for Experiments 3a and 3b is available here: <https://osf.io/734m8/> and <https://osf.io/ygt6b/>

### **Acknowledgments**

This work was supported by NEI EY017001 (JMW). The authors thank Hayden Schill and Makaela Nartker for assistance with data collection, Joey Feffer, Erin Smith, and Jungyeon Park for assistance with analysis, and Matthew Cain for helpful discussions.

### **References**

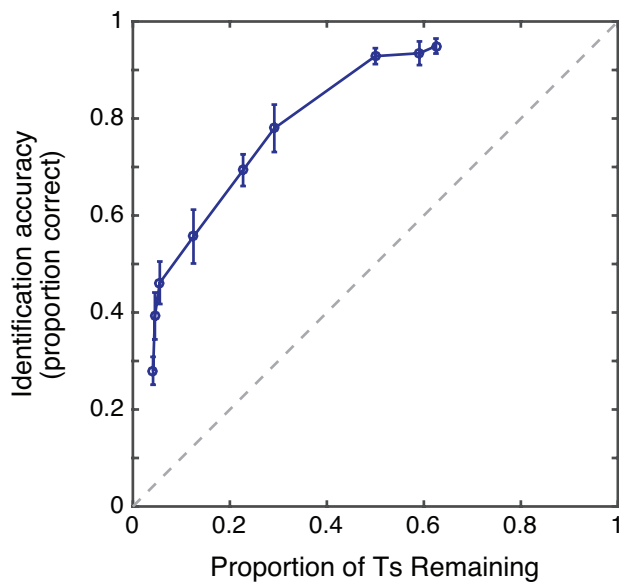
- Baldauf, D., & Deubel, H. (2008). Properties of attentional selection during the preparation of sequential saccades. *Experimental Brain Research*, 184(3), 411–425. <http://doi.org/10.1007/s00221-007-1114-x>
- Baldauf, D., & Deubel, H. (2010). Attentional landscapes in reaching and grasping. *Vision Research*, 50(11), 999–1013. <http://doi.org/10.1016/j.visres.2010.02.008>
- Berbaum, K. S. (2012). Satisfaction of search experiments in advanced imaging. *Human Vision and Electronic Imaging XVII*, 8291(February 2012), 82910V. <http://doi.org/10.1117/12.916461>
- Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software*, 31(10). <http://doi.org/10.18637/jss.v031.i10>
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, 11(5), 4–4. <http://doi.org/10.1167/11.5.4>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <http://doi.org/10.1163/156856897X00357>
- Cain, M. S., Vul, E., Clark, K., & Mitroff, S. R. (2012). A Bayesian Optimal Foraging Model of Human Visual Search. *Psychological Science*, 23(9), 1047–1054. <http://doi.org/10.1177/0956797612440460>
- Callahan-Flintoft, C., & Wyble, B. (2017). Measuring the speed of attentional selection for two features concurrently from a single object reveals a foveal speed advantage for color [Abstract]. *Vision Sciences Society*, 17(1335). <http://doi.org/10.1167/17.10.1335>
- Carlson, T. A., Hogendoorn, H., & Verstraten, F. A. J. (2006). The speed of visual attention: What time is it? *Journal of Vision*, 6(2006), 1406–1411. <http://doi.org/10.1167/6.12.6>
- Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences*, 101(35), 13086–13090. <http://doi.org/10.1073/pnas.0305329101>
- Charnov, E. L. (1976). Optimal foraging: The marginal value theorem. *Theoretical Population Biology*, 9(2), 129–136.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617. <http://doi.org/10.3758/BF03195489>
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5), 1–36. <http://doi.org/10.1167/11.5.14>
- Ehinger, K. A., & Wolfe, J. M. (2016). When is it time to move to the next map? Optimal foraging in



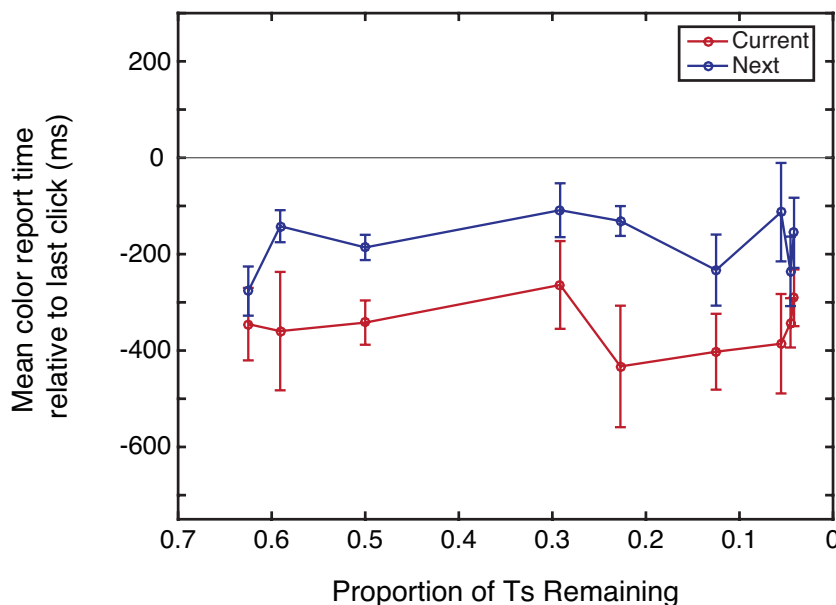
- guided visual search. *Attention, Perception, and Psychophysics*, 78(7), 2135–2151.  
<http://doi.org/10.3758/s13414-016-1128-1>
- Fleck, M. S., Samei, E., & Mitroff, S. R. (2010). Generalized “Satisfaction of Search”: Adverse Influences on Dual-Target Search Accuracy. *Journal of Experimental Psychology: Applied*, 16(1), 60–71. <http://doi.org/10.1037/a0018629>
- Gersch, T. M., Kowler, E., Schnitzer, B. S., & Doshier, B. A. (2009). Attention during sequences of saccades along marked and memorized paths. *Vision Research*, 49(10), 1256–1266.  
<http://doi.org/10.1016/j.visres.2007.10.030>
- Godijn, R., & Theeuwes, J. (2003). Parallel Allocation of Attention Prior to the Execution of Saccade Sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 29(5), 882–896. <http://doi.org/10.1037/0096-1523.29.5.882>
- Horowitz, T. S., & Thornton, I. M. (2008). Objects or locations in vision for action? Evidence from the MILO task. *Visual Cognition*, 16(4), 486–513. <http://doi.org/10.1080/13506280601087356>
- Hunt, A. R., & Cavanagh, P. (2009). Looking ahead: The perceived direction of gaze shifts before the eyes move. *Journal of Vision*, 9(9), 1–7. <http://doi.org/10.1167/9.9.1>
- Kleiner, M., Brainard, D., & Pelli, D. G. (2007). What’s new in Psychtoolbox-3? *Perception 36 ECVF Abstract Supplement*, 36(14), 1.
- Kosovicheva, A., & Bex, P. (2016). Reading rainbows: Measuring the dynamics of word processing [Abstract]. *Journal of Vision*, 16(12), 1314. <http://doi.org/10.1167/16.12.1314>
- Kristjánsson, Á., Jóhannesson, Ó. I., & Thornton, I. M. (2014). Common attentional constraints in visual foraging. *PLoS ONE*, 9(6). <http://doi.org/10.1371/journal.pone.0100752>
- Kristjánsson, T., Thornton, I. M., Chetverikov, A., & Kristjánsson, Á. (2020). Dynamics of visual attention revealed in foraging tasks. *Cognition*, 194. <http://doi.org/10.1016/j.cognition.2019.104032>
- Kwak, H., Dagenbach, D., & Egeth, H. (1991). Further evidence for a time-independent shift of the focus of attention. *Perception & Psychophysics*, 49(5), 473–480.
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, 17(8), 391–400.  
<http://doi.org/10.1016/j.tics.2013.06.006>
- McPeck, R. M., Skavenski, A. A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision Research*, 40(18), 2499–2516. [http://doi.org/10.1016/S0042-6989\(00\)00102-4](http://doi.org/10.1016/S0042-6989(00)00102-4)
- Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research*, 51(13), 1526–1537.  
<http://doi.org/10.1016/j.visres.2010.09.003>
- Pashler, H. (1991). Shifting visual attention and selecting motor responses: Distinct attentional

- mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, 17(4), 1023–1040. <http://doi.org/10.1037/0096-1523.17.4.1023>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <http://doi.org/10.1163/156856897X00366>
- Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*, 25(2), 137–142. <http://doi.org/10.3758/BF03204486>
- Sung, K. (2008). Serial and parallel attentive visual searches: Evidence from cumulative distribution functions of response times. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1372–1388. <http://doi.org/10.1037/a0011852>
- Thornton, I. M., & Horowitz, T. S. (2004). The multi-item localization (MILO) task: Measuring the spatiotemporal context of vision for action. *Perception and Psychophysics*, 66(1), 38–50. <http://doi.org/10.3758/BF03194859>
- Tuddenham, W. J. (1962). Visual search, image organization, and reader error in roentgen diagnosis: Studies of the psychophysiology of roentgen image perception. *Radiology*, 78, 694–704. <http://doi.org/10.1148/78.5.694>
- Wolfe, J. M. (2012). Saved by a Log: How Do Humans Perform Hybrid Visual and Memory Search? *Psychological Science*, 23(7), 698–703. <http://doi.org/10.1177/0956797612443968>
- Wolfe, J. M. (2013). When is it time to move to the next raspberry bush? Foraging rules in human visual search. *Journal of Vision*, 13(3), 1–17. <http://doi.org/10.1167/13.3.10>
- Wolfe, J. M. (2018). Visual Search. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (4th Ed.). Hoboken, NJ, USA: John Wiley & Sons, Inc. <http://doi.org/10.1002/9781119170174.epcn213>
- Wolfe, J. M., Aizenman, A. M., Boettcher, S. E. P., & Cain, M. S. (2016). Hybrid foraging search: Searching for multiple instances of multiple types of target. *Vision Research*, 119, 50–59. <http://doi.org/10.1016/j.visres.2015.12.006>
- Wu, C., & Wolfe, J. M. (2018). A new multiple object awareness paradigm shows that imperfect knowledge of object location is still knowledge. *Current Biology*, 28(21), 3430–3434. <http://doi.org/10.1016/j.cub.2018.08.042>
- Wundt, W. (1883). *Philosophische Studien*. Leipzig: Wilhelm Engelmann.
- Zhang, J., Gong, X., Fougner, D., & Wolfe, J. M. (2015). Using the past to anticipate the future in human foraging behavior. *Vision Research*, 111(Part A), 66–74. <http://doi.org/10.1016/j.visres.2015.04.003>

----- **Supplemental Figures** -----

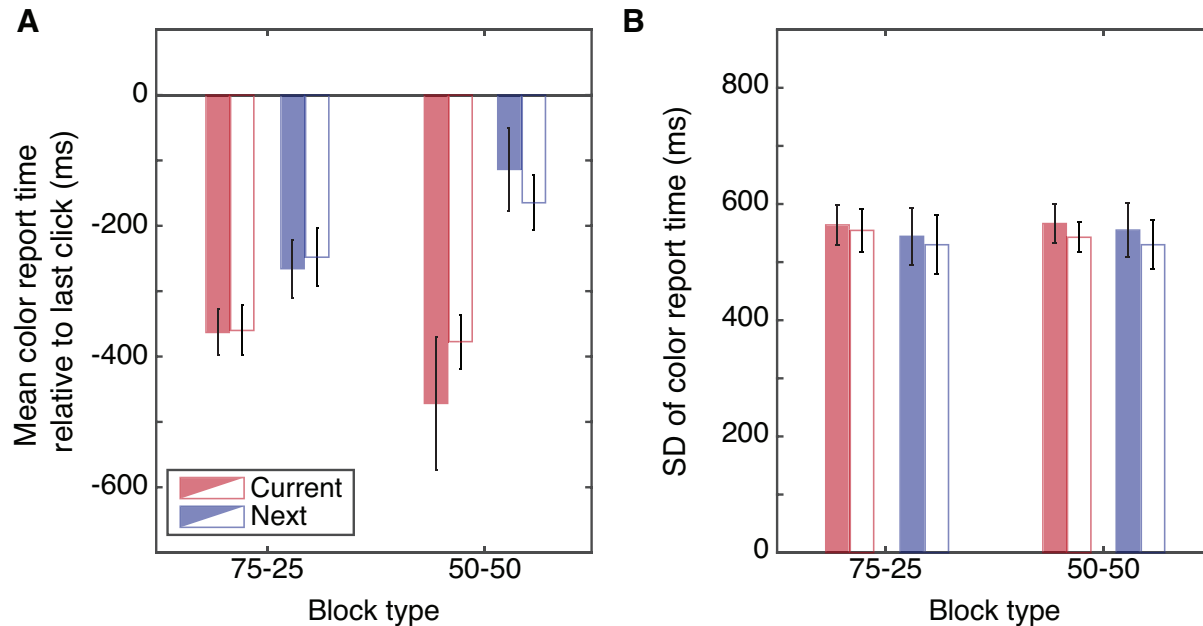


**Figure S1.** Identification accuracy on the placeholder screen in Experiment 1 (applies to Next trials only) as a function of the proportion of targets remaining in the display. Gray dashed line shows the chance level of performance. Observers were consistently above chance in identifying Ts on the placeholder screen across all Target Density conditions. As expected, observers' performance in identifying the locations of future targets was lower when targets were scarcer (i.e., low proportion of Ts remaining). Error bars represent  $\pm 1$  SEM.

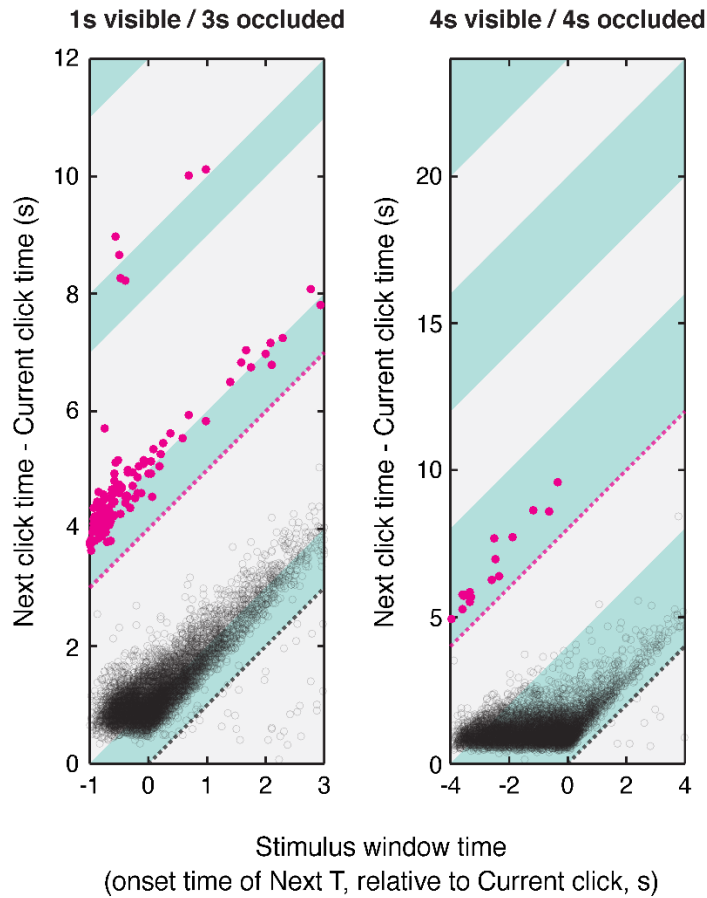


**Figure S2. Analysis of Experiment 1 including placeholder-inaccurate trials.** Mean time corresponding to color reports, with *all* trials included in the analysis, for the Current and Next conditions (red and blue, respectively), plotted as function of the proportion of Ts remaining on the

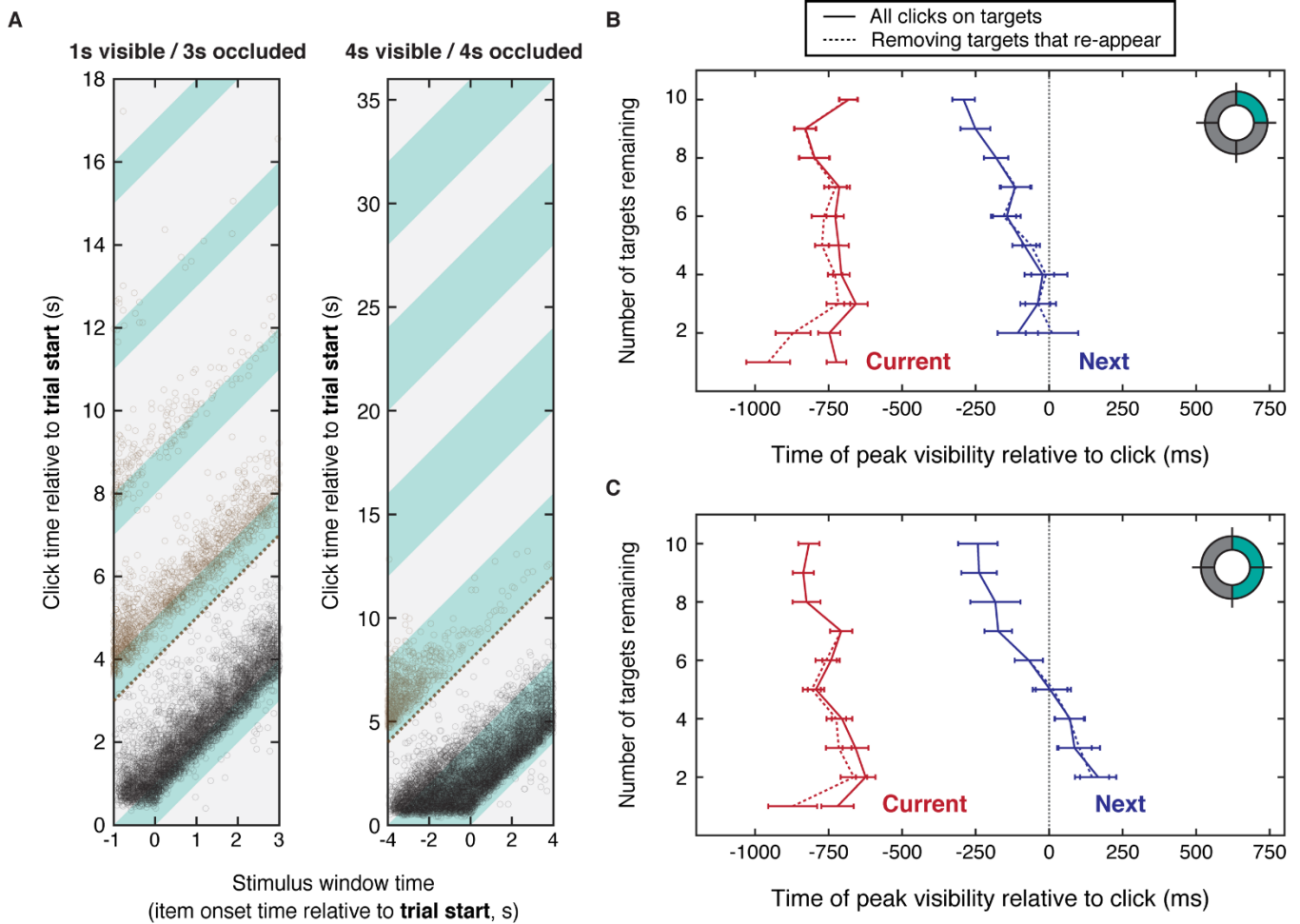
screen. These results are similar to those shown in Figure 2C, which shows only the location-accurate trials. In other words, the results are unaffected by the exclusion of location-inaccurate trials; in both analyses, observers report later colors for the Next item compared to the Current item, and this does not change with the proportion of Ts remaining. Error bars represent  $\pm 1$  SEM.



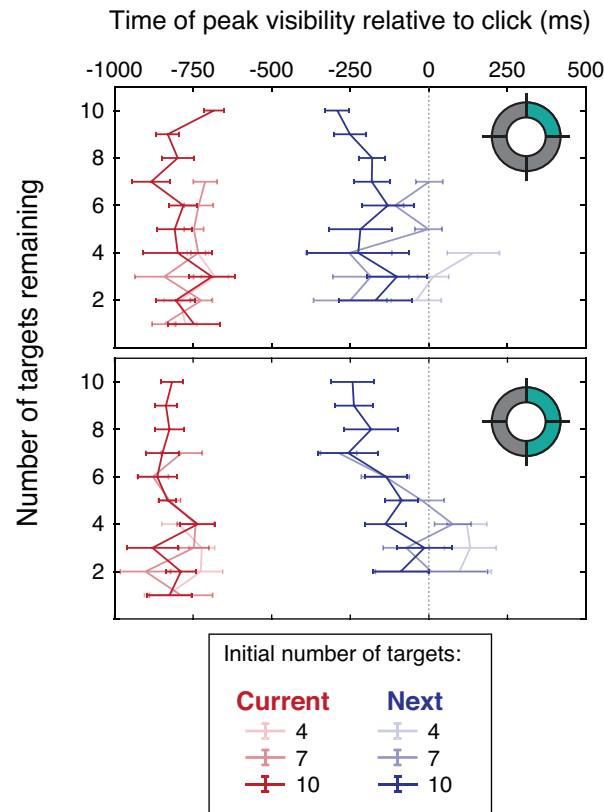
**Figure S3. Results from additional experiment (Supplemental Note 2) varying relative frequency of Current vs. Next trials.** The relative frequency of Current vs. Next trials was varied across two different Block Types (75-25% vs. 50-50% probability of Current vs. Next Trials). The left and right panels show the means and SDs of observers' color reports (A and B, respectively). Color reports are shown separately for each Block Type and for Current vs. Next trials (red and blue, respectively). Filled bars represent results with all observers included, and empty bars show results with one outlier removed, which affects mean values in the 50-50% block (see Supplemental Note 2 for full description of the experiment and statistical analyses). Neither the mean reported color nor the SD of the color distribution was affected by changing the frequency with which observers were asked to report the colors of Current versus Next items. This indicates that the results observed in Experiment 1 are unlikely to be influenced by changes in task demands. Error bars represent  $\pm 1$  SEM.



**Figure S4.** Analysis showing target visibility in-between clicks for Experiment 3b. This plot is similar to Figure 8 in the main paper, showing clicks in Exp 3b instead, where items are randomly shuffled between locations on each click. The y-axis shows the time interval between the Current and Next click, and the x-axis shows the time point within the stimulus visibility cycle corresponding to the Next clicked item (i.e., the onset time of the Next target, relative to the Current click). The black open circles represent clicks that occurred before the Next target re-appeared during the trial (i.e., the target was visible no more than once). Note that in this experiment, in contrast to Exp. 3a, there are very few clicks below the black dotted line. These are instances in which the target was fully occluded in-between clicks, and therefore represent lucky guesses (i.e., the observer was given no information about the target, but correctly guessed the location of a T). Data points above the magenta dotted line (filled magenta circles) are clicks that occurred *after* the target re-appeared during the trial (examples shown in panel A of Figure 8). These account for 1.03% of all clicks in Experiment 3b. As in Experiment 3a, observers in Experiment 3b rarely wait for targets to re-appear in-between clicks, and we can conclude that items tend to be found during the interval closest to the click.



**Figure S5.** Analysis of target visibility for all clicks in Experiment 3a. (A) Scatterplot of all clicks, similar to Figure 8 in the main paper, with the x- and y-axes referenced to the start of the trial, instead of the preceding click. In other words, the y-axis shows the time interval between the start of the trial and each click, and the x-axis shows the time point within the stimulus visibility cycle corresponding to the start of the trial (i.e., the onset time of the item, relative to the start of the trial). Negative values represent targets that are already visible when the trial starts. The black open circles represent targets that were visible no more than once prior to being clicked. In these instances, there is only one visible window, and therefore the target could only have been found during that time. Data points above the brown dotted line (brown circles) are clicks that occurred after the target was visible multiple times during the trial (and therefore there are multiple intervals in which the target could have been found). These account for 17.3% of all clicks in Experiment 3b (24.3% and 10.4% of all clicks in the 1s-on/3s-off and the 4s-on/4s-off conditions, respectively). (B & C) We re-analyzed the visibility distributions, removing the clicks represented by the brown circles in panel A (i.e., cases in which there is ambiguity about when the target could have been found). Dashed lines represent the re-analyzed data, where time of peak visibility is plotted as a function of the number of targets remaining, separately for the 1s-on/3s-off (B) and the 4s-on/4s-off conditions (C). Solid lines represent the data with all clicks included in the analysis, re-plotted from Figure 6 in the main paper. Red and blue lines correspond to the Current and Next analyses, respectively. These plots indicate that we observe similar results if we remove cases in which there is ambiguity about when the target could have been found (i.e., removing all clicks in which the target is visible for more than one interval). Error bars represent  $\pm 1$  SEM.



**Figure S6.** Analysis of results from Experiment 3a, separated by set size (compare to Figure 6 in the main text). Time of peak visibility relative to the time of the click on the current target, for the 1s-on/3s-off condition (upper panel), and the 4s-on/4s-off condition (lower panel). Red and blue lines show the peak finding times for the Current and Next analyses, respectively. These results show that, as expected, set size doesn't affect the finding time (peak visibility) of the Current target relative to the current click, as these values represent the finding-to-click time. However, set size may contribute to the progressively later finding times of the next target, as the number of remaining targets drops. When the display contains 4 targets initially, next items are found somewhat later (compared to displays that started with a larger number of targets). It is possible that the instantaneous target-to-distractor ratio contributes to this difference: when 4 targets remain, a display that started with 10Ts will have a higher target density (4/18 or 22%) compared to a display that started with 4Ts (4/24 or 16.7%).

#### ----- Supplemental Movies -----

The videos below summarize the results from Experiment 2a using two dimensional maps. Two types of spatial maps were calculated: one showing the fixation onsets (arrival time, similar to Figure 4A in the main paper), and one showing the density of gaze position at each time point (similar to Figure 4B). Both were calculated in 20 ms increments from -1300 ms to +1300 ms around the time of the click. The spatial locations of Target<sub>N</sub> and Target<sub>N+1</sub> were aligned across clicks by individually scaling and rotating them, such that the location of Target<sub>N</sub> is at x = 0, y = 0 (represented by the red square) and the location of Target<sub>N+1</sub> is at x = 1, y = 0 (represented by the white square). As a result, all the x- and y-positions shown are in arbitrary units. In the maps showing fixation onsets, individual "X"s represent all new fixations (across observers) within  $\pm 10$  ms of the time shown at the top of the plot. As these represent individual fixation onsets, any given "X" is shown for only one frame.

Similar to the data shown in Figure 4, both fixation onsets and fixation density switch from Target<sub>N</sub> to Target<sub>N+1</sub> close to the time of the click on target N (0 ms), indicating that observers tend to fixate the next item close to the time of the current click. As in Figure 4, fixation onsets lead slightly ahead of the density maps.

[Movie S1](#). Map of fixation onsets and density together.

[Movie S2](#). Map of fixation onsets only.

[Movie S3](#). Map of fixation density only.

## ----- Supplemental Notes -----

### Supplemental Note 1

Analysis of data across all trials in Experiment 1 (location accurate and location inaccurate)

Trials with inaccurate identification responses on the placeholder screen (collected on Next trials only) were removed from the analysis in Experiment 1, as we expected these trials to contribute additional variability to observers' color reports. We separately analyzed observers' color responses, including all trials in the analysis (location accurate and inaccurate). The inclusion of location-inaccurate trials significantly increased the SD of the response distribution in the Next trials, from 526 ms to 638 ms,  $t(11) = 14.42$ ,  $p < .001$ . However, the difference in SD between Current and Next trials (573 vs. 638 ms) did not reach significance,  $t(11) = 1.53$ ,  $p = 0.16$ . Even with this added variability, the SDs of the response distributions on Next trials were still narrower than those expected by random guessing for each of the twelve observers (all  $p$ -values  $< .001$ ; permutation test;  $\alpha_B = .004$ ).

As expected, the inclusion of these trials did not shift the mean of the response distributions. Figure S2 shows mean reported time with all trials included in the analysis (compare to Figure 2C). Similar to the results reported in Experiment 1, a 2 (Current vs Next)  $\times$  9 (Target Density) ANOVA showed a significant main effect of Trial Type ( $F(1,11) = 11.15$ ,  $p = .007$ ,  $\eta_p^2 = .50$ ), with observers reporting earlier colors in Current trials compared to Next trials (-330 ms vs -174 ms, respectively). Neither the main effect of Target Density ( $F(8,88) = 0.93$ ,  $p = 0.50$ ,  $\eta_p^2 = .08$ ), nor the Target Density  $\times$  Trial Type interaction ( $F(8,88) = 0.5$ ,  $p = 0.85$ ,  $\eta_p^2 = .04$ ) reached significance.

### Supplemental Note 2

Control experiment for Experiment 1: Varying the relative frequency of Current and Next trials

In Experiment 1, there was an even (50-50%) distribution of Current and Next trials. One possible account of the results in Experiment 1 is that observers may intentionally adjust their strategy, knowing that they would need to frequently report information about future targets. To evaluate this, we ran a separate experiment, varying the relative frequency of Current and Next trials in separate blocks (referred to as the Block Type manipulation). In one condition, the distribution of Current and Next trials was even (50-50%), and in the other condition, there were fewer Next trials compared to Current trials (25% and 75%, respectively).

## Method

### Observers

Twelve observers (5 female, mean age: 30.6, range: 20-54) participated in the experiment.



## Stimuli and Procedure

The experiment was an earlier variant of the design of Experiment 1 and therefore had several minor methodological differences from Experiment 1, in addition to the block type manipulation. Each search display consisted of 30 items arranged into a 5 row  $\times$  6 column screen-centered grid. Item locations were initially distributed across an area  $12.7^\circ$  wide  $\times$   $12.2^\circ$  high, and then had a random spatial jitter added to each item, between  $-0.86^\circ$  and  $+0.86^\circ$  in the horizontal direction and between  $-1.11^\circ$  and  $+1.11^\circ$  in the vertical direction. Each shape was  $0.8^\circ$  in height and width (line width of  $0.24^\circ$ ). The number of targets at the onset of each trial ( $N$ ) was randomly selected from between 4 and 9, and the trial exited unpredictably after a random number of targets, selected between 2 and  $N-1$  had been clicked.

On Next trials, observers were shown a placeholder screen with boxes covering only the locations of the remaining targets, and observers indicated which one of these targets they intended to click next. If observers knew where they were going next, they were instructed to left-click the gray box. Otherwise, if they had no knowledge of the locations of the remaining Ts, they were instructed to right-click on one of the boxes. Only trials in which observers knew where they were going next were included in the analysis (see Analysis). Following the placeholder response, the trial proceeded to the color response screen. As in Experiment 1, observers then indicated the color of the item on a response palette (inner and outer radii of the annulus were  $3.97^\circ$  and  $6.86^\circ$ , respectively). On Current trials, the experiment proceeded immediately to the color selection screen after the last T was clicked. Confidence ratings were collected separately, following the color responses, using a 1-5 Likert scale, with 5 being the most confident. A confidence rating screen was shown immediately after observers indicated their color selection. The text “How certain were you about the color of the target?” was shown at the top of the display, and below it, observers were shown a row of 5 gray circles. The labels “I guessed”, “Somewhat sure”, and “I am certain” appeared below circles 1, 3, and 5, respectively. Observers clicked on the circle that best corresponded to their confidence level. As in Experiment 1, after making their response, observers were shown a feedback screen listing the number of targets collected, the total trial duration, and the average time per item on the previous trial. Observers then advanced to the next trial by clicking the mouse button when ready.

Observers completed two blocks of trials: one with a 75% probability of Current trials (referred to as the 75-25% block), and the other with a 50% probability of Current trials (the 50-50% block). Five observers completed the 50-50% block first and 7 observers completed the 75-25% block first. Each block consisted of 130 trials. Before starting the experiment, to familiarize themselves with the task, observers completed a practice block consisting of 40 trials, which had a 60% probability of Current trials. Aside from the differences noted above, all other experimental procedures were identical to Experiment 1.

## Analysis

On Next trials, all the box options on the placeholder screen were Ts. Therefore, we used observers’ self-reports (right-click or left-click) on the gray boxes, rather than their identification accuracy, as an indicator of whether they knew where they were going next. For Next trials, only location-certain trials (i.e., trials with a left-click on the gray boxes) were included in the analysis. This accounted for 60.7% of all Next trials. This proportion was not significantly different between the two Block Types (62.8% vs 58.6%;  $t(11) = 0.76$ ,  $p = .46$ ), and comparable to the proportion of location-accurate Next trials in Experiment 1 (66.4%). In this experiment, we were primarily interested in the effects of the Block Type manipulation. Therefore, we analyzed the effects of the two Block Types (50/50 and 75/25) and two Trial Types (Current and Next) on both the means and the SDs of the response distributions using separate 2 (Block Type)  $\times$  2 (Trial Type) repeated-measures ANOVAs. We note that, for one observer, the overall difference between mean color report on Current and Next trials was  $> 2.5$  SDs from the

group average difference (948 vs 228 ms). We report the ANOVA results with the inclusion of this observer, though we note the results are similar when we exclude this observer from the analysis. For illustrative purposes, Figure S3 shows results both with and without this observer.

## Results

Figure S3 shows the means and SDs of the response distributions, for each Trial Type (Current and Next) and Block Type (50/50 and 75/25). The ANOVA on the means of the response distributions showed a significant main effect of Trial Type (Current vs. Next):  $F(1,11) = 8.08, p = .016, \eta_p^2 = .42$ , with earlier responses in Current trials compared to Next trials (-417 vs -190 ms, respectively), which replicated the results of Experiment 1. However, there was no significant main effect of Block Type ( $F(1,11) = .55, p = .48, \eta_p^2 = .05$ ), and no significant Trial Type x Block Type Interaction ( $F(1,11) = 2.31, p = .16, \eta_p^2 = .17$ ).

Similar to Experiment 1, the ANOVA on the SDs of the response distributions showed no significant main effect of Trial Type (Current vs. Next; 565 vs 550 ms, respectively):  $F(1,11) = .17, p = .69, \eta_p^2 = .02$ . Neither the main effect of Block Type ( $F(1,11) = .10, p = .76, \eta_p^2 = .009$ ), nor the Trial Type x Block Type Interaction ( $F(1,11) = .04, p = .85, \eta_p^2 = .003$ ) were significant. Together, these results indicate a minimal influence of the frequency of Next trials on either the mean or the variability of observers' color responses.

### Supplemental Note 3 Full ANOVA results for Experiment 3

In Experiment 3, we compared the peaks of the visibility distributions between conditions, using a 2 (Experiment: Shuffle vs No-shuffle) x 2 (Visibility Condition) x 2 (Target Type: Current or Next) x 9 (Targets Remaining) mixed-model ANOVA, with Experiment as a between-subjects factor and the remaining factors as within-subjects factors. The results of the ANOVA are listed below:

|                               | Num.<br>DF | Denom.<br>DF | F Ratio | p-value | $\eta_p^2$ |
|-------------------------------|------------|--------------|---------|---------|------------|
| Experiment                    | 1          | 22           | 165.16  | <.0001  | 0.88       |
| Visibility                    | 1          | 22           | 1.33    | 0.2604  | 0.06       |
| Trial Type                    | 1          | 22           | 1142.29 | <.0001  | 0.98       |
| Targets Remaining (TR)        | 8          | 176          | 16.42   | <.0001  | 0.43       |
| Exp × Visibility              | 1          | 22           | 0.79    | 0.3827  | 0.03       |
| Exp × Trial Type              | 1          | 22           | 58.87   | <.0001  | 0.73       |
| Exp × TR                      | 8          | 176          | 5.11    | <.0001  | 0.19       |
| Visibility × Trial Type       | 1          | 22           | 7.90    | 0.0102  | 0.26       |
| Visibility × TR               | 8          | 176          | 6.39    | <.0001  | 0.23       |
| Trial Type × TR               | 8          | 176          | 4.00    | 0.0002  | 0.15       |
| Exp × Visibility × Trial Type | 1          | 22           | 0.003   | 0.9601  | <.001      |
| Exp × Visibility × TR         | 8          | 176          | 0.43    | 0.902   | 0.02       |
| Exp × Trial Type × TR         | 8          | 176          | 1.71    | 0.0983  | 0.07       |
| Visibility × Trial Type × TR  | 8          | 176          | 1.56    | 0.14    | 0.07       |

|                                    |   |     |      |        |      |
|------------------------------------|---|-----|------|--------|------|
| Exp × Visibility × Trial Type × TR | 8 | 176 | 1.44 | 0.1826 | 0.06 |
|------------------------------------|---|-----|------|--------|------|

### Supplemental Note 4

Simulations for Experiment 3 – Relationship between visibility distribution and target finding time

#### Simulation Method

We used simulations of the task in Experiment 3 to evaluate the degree to which the peaks of the visibility distributions reflect the times that targets were found. Our general approach was to create computer simulations that closely reproduced the task and to then find parameters that closely matched the observed data. The simulations made few assumptions about the search process; instead, we systematically varied seven parameters and compared the simulated visibility distributions in each case to the observed visibility (Figure 7B). Next, we compared the underlying finding times from the simulation to the visibility distributions observed in the experiment (Figure 7C).

Figure 7A in the main paper outlines the general simulation procedure and the table below summarizes all simulation parameters. We reproduced trials in Experiment 3, matching the number of Targets (4, 7, or 10) and distractors (20, 17, or 14), and generated variables storing the identities (T or L) and  $x$  and  $y$  positions of the items in the display. As in the experiment, each item was assigned a randomly selected phase for visibility/occlusion, and we generated visibility timelines for each item over the whole trial. Unoccluded items were sequentially “checked” and collected according to the following procedure. Each simulated trial started with the program randomly selecting an unoccluded item to visit (i.e., to check whether it was a T or an L). A fixed delay was associated with checking the item (*Check Duration*), and these delays were added together for consecutive items until a T was found. If the checked item was a T, a “finding time” was assigned based on the end of this search interval. Then, collection of the target was simulated by adding a motor response delay, based on the combination of the distance of the most recently clicked item and the *Response Speed* parameter. This simulated a ‘cursor’ that moved from item to item. (For the first click within a trial, this distance was calculated from a random location.) Future targets were checked based on one of two *Checking Rules* – either randomly from the array, without regard to spatial location, or weighted by the distance from the previously checked item (Gaussian profile with  $\sigma = 200$  pixels, or  $5.36^\circ$ ). In both cases, checked items were always chosen from the set of unoccluded and uncollected items. As in the experiment, items had to be visible in order to be found, but could be clicked while they were occluded. An additional *Revisit Rule* parameter determined whether the same item could be checked repeatedly, or no more than once.

Two additional constraints were common across all simulations. First, the search had to occur serially, meaning that multiple items could not be ‘checked’ simultaneously. Second, collection also occurred serially, such that a motor response to a target could not begin until the previous one was completed. However, to simulate observers’ ability to plan ahead, additional targets could be checked during the motor response intervals for previous targets. This allowed the buildup of a queue of Ts that had been found, but not collected yet. Additional items were checked provided the number of items in the queue did not exceed  $N$  (the *Planning Limit* parameter). For example, if the Planning Limit for a given simulation is 2, and there are 2 items that have been found, but not collected, checking is withheld until one of those items has been collected. Importantly, a Planning Limit of 0 means that the search for the next item cannot begin until the current item is collected (i.e., there is no ability to plan ahead). The *Clicking Rule* parameter determined whether items in the queue were collected based on the order in which they were found, or in order of the item physically closest to the most recent click. For simplicity,

once a motor response was initiated to a target, it could not be changed to a different target. Finally, each trial terminated once either all Ts had been collected, or once the time since the previously found item exceeded the *Quit Time* parameter (but only if at least two Ts had been collected, matching the original experiment).

| Parameter      | Description   | Values tested                  |
|----------------|---|--------------------------------|
| Check duration | Time spent checking (“visiting”) a given item to determine whether it is a T or L (msec)  | 50 – 290<br>in steps of 30     |
| Response speed | Speed of motor response in collecting the target (pixels/sec). Motor response duration was calculated from: $\text{Distance from previous target} / \text{Response Speed}$ .  | 400 – 800<br>in steps of 50    |
| Quit time      | Trial is exited if the elapsed time since previously found item exceeds N (msec)  | 1000 – 6000<br>in steps of 500 |
| Revisit rule   | Items are visited a maximum of one time (0) or can be checked multiple times (1)  | 0,1                            |
| Checking rule  | Next checked item is chosen randomly from the set (0), or weighted by distance from the previously checked item (1)   | 0,1                            |
| Clicking rule  | Items are clicked based on the order in which they were found (0), or based on the nearest item to the most recent click (1)  | 0,1                            |
| Planning limit | Maximum size of click queue. A new item is visited provided the number of items that have been found, but not clicked yet does not exceed N. A value of 0 means that the search for the next item cannot begin until the current item is collected. | 0,1,2,3,4                      |

The values listed in the table above produced 35,640 unique combinations of parameters, which were tested on separate iterations of the program. Each iteration terminated after 1000 targets had been collected. This procedure was performed separately for the 1/3 and 4/4 sec conditions (approximately  $7.1 \times 10^7$  targets in total). For each iteration, we analyzed the data as in the original experiment, and calculated the sum-of-squared errors relative to the observed visibility distributions. The observed visibility distributions were calculated from all clicks within a condition, across observers, separately for the Targets Remaining for the Current and Next analyses. On each iteration, the sum-of-squared errors was summed across all of the Targets Remaining and the Current and Next analyses.

### Simulation Results

The underlying finding times from the best-fitting simulations are shown in Figure 7 and described in the main body of the paper. In addition, the figure below summarizes variation in the sum-of-squared errors (SSEs) for each of these parameters (collapsing across the remaining parameters). The x-axis shows the value of the simulation parameter (see above table for description), and the y-axis shows the sum-of-squared errors (summed across Targets Remaining and Current/Next analyses). Purple and green lines represent the 1s-on / 3s-off and 4s-on / 4s-off conditions, respectively. Error bars represent  $\pm 1$  SEM.

