

## Learning Where to Look for a Hidden Target

---

LEANNE CHUKOSKIE,<sup>\*</sup> JOSEPH SNIDER,<sup>\*</sup>  
MICHAEL C. MOZER,<sup>\*†‡</sup> RICHARD J. KRAUZLIS,<sup>§</sup> AND  
TERRENCE J. SEJNOWSKI<sup>\*||#\*\*\*††</sup>

Survival depends on successfully foraging for food, for which evolution has selected diverse behaviors in different species. Humans forage not only for food, but also for information. We decide where to look over 170,000 times per day, approximately three times per wakeful second. The frequency of these saccadic eye movements belies the complexity underlying each individual choice. Experience factors into the choice of where to look and can be invoked to rapidly redirect gaze in a context- and task-appropriate manner. However, remarkably little is known about how individuals learn to direct their gaze given the current context and task. We designed a task in which participants search a novel scene for a target whose location was drawn stochastically on each trial from a fixed prior distribution. The target was invisible on a blank screen, and the participants were rewarded when they fixated the hidden target location. In just a few trials, participants rapidly found the hidden targets by looking near previously rewarded locations and avoiding previously unrewarded locations. Learning trajectories were well characterized by a simple reinforcement-learning (RL) model that maintained and continu-

---

<sup>\*</sup>Institute for Neural Computation and <sup>\*\*</sup>Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093; <sup>†</sup>Department of Computer Science and <sup>‡</sup>Institute of Cognitive Science, University of Colorado, Boulder, CO 80309; <sup>§</sup>Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD 20892; and <sup>||</sup>Howard Hughes Medical Institute and <sup>#</sup>Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037. <sup>††</sup>To whom correspondence should be addressed. E-mail: terry@salk.edu.

ally updated a reward map of locations. The RL model made further predictions concerning sensitivity to recent experience that were confirmed by the data. The asymptotic performance of both the participants and the RL model approached optimal performance characterized by an ideal-observer theory. These two complementary levels of explanation show how experience in a novel environment drives visual search in humans and may extend to other forms of search such as animal foraging.

**T**he influence of evolution can be seen in foraging behaviors, which have been studied in behavioral ecology. Economic models of foraging assume that decisions are made to maximize payoff and minimize energy expenditure. For example, a bee setting off in search of flowers that are in bloom may travel kilometers to find food sources. Seeking information about an environment is an important part of foraging. Bees need to identify objects at a distance that are associated with food sources. Humans are also experts at searching for items in the world, and in learning how to find them. This study explores the problem of how humans learn where to look in the context of animal foraging.

Our daily activities depend on successful search strategies for finding objects in our environment. Visual search is ubiquitous in routine tasks: finding one's car in a parking lot, house keys on a cluttered desk, or the button you wish to click on a computer interface. When searching common scene contexts for a target object, individuals rapidly glean information about where targets are typically located (Potter, 1975; Itti and Koch, 2000; Neider and Zelinsky, 2006; Oliva and Torralba, 2006; Torralba et al., 2006; Rayner et al., 2009; Castelhana and Heaven, 2010, 2011; Vö and Henderson, 2010). This ability to use the "gist" of an image (Oliva and Torralba, 2006; Torralba et al., 2006) enables individuals to perform flexibly and efficiently in familiar environments. Add to that the predictable sequence of eye movements that occurs when someone is engaged in a manual task (Hayhoe and Ballard, 2005) and it becomes clear that despite the large body of research on how image salience guides gaze (Itti and Koch, 2000; Parkhurst and Niebur, 2003), learned spatial associations are perhaps just as important for effectively engaging our visual environment (Chun and Jiang, 1998; Hayhoe and Ballard, 2005; Tatler and Vincent, 2009). Surprisingly, however, little research has been directed to how individuals learn to direct gaze in a context- and task-appropriate manner in novel environments.

Research relevant to learning where to look comes from the literature on eye movements, rewards, and their expected value. Like all motor behavior, saccades are influenced by reward, occurring at shorter latency for more valued targets (Milstein and Dorris, 2007). In fact, finding something you seek may be intrinsically rewarding (Xu-Wilson et al., 2009).

Refining the well-known canonical “main sequence” relationship between saccade amplitude and velocity, the value of a saccade target can alter details of the motor plan executed, either speeding or slowing the saccade itself depending upon the value of that target for the subject (Shadmehr, 2010; Shadmehr et al., 2010). This result is especially interesting in light of the research indicating that the low-level stimulus features, which have an expected distribution of attracting fixations (Reinagel and Zador, 1999), are different (Tatler et al., 2006) and perhaps also differently valuable (Açık et al., 2010) depending on their distance from the current fixation location. Taken together these results underscore the complex interplay of external and internal information in guiding eye movement choice.

Two early foundational studies from Buswell (1935) and Yarbus (1967) foreshadowed modern concepts of a priority or salience map by showing that some portions of an image are fixated with greater likelihood than others. Both researchers also provided early evidence that this priority map effectively changes depending on the type of information sought. Yarbus observed that the patterns of gaze that followed different scene-based questions or tasks given to the observer were quite distinct, suggesting that the observer knew where to find information in the scene to answer the question and looked specifically to areas containing that information when it was needed. Henderson and coworkers (Castelhano et al., 2009) have replicated this result for the different tasks of visual search and image memorization. However, Wolfe and coworkers (Greene et al., 2012), using a slightly different question and task paradigm, failed to find evidence that saccade patterns were predictive of specific mental states. Regardless of specific replications of Yarbus’s demonstration, it is clear that scene gist—context-specific information about where objects are typically found—emerges very quickly and guides target search of a scene with a known context (Torralba et al., 2006). For example, when shown a street scene, an observer would immediately know where to look for street signs, cars, and pedestrians (Fig. 14.1A).

Castelhano and Heaven (2011) have also shown that in addition to scene gist itself, learned spatial associations guide eye movements during search. Subjects use these learned associations as well as other context-based experience, such as stimulus probability, and past rewards and penalties (Geng and Behrmann, 2005; Stritzke and Trommershäuser, 2007; Schütz et al., 2012) to hone the aim of a saccadic eye movement. A recent review and commentary from Wolfe et al. (2011) explores the notion of “semantic” guidance in complex, naturalistic scenes as providing knowledge of the probability of finding a known object in a particular part of a scene. This perspective relates work on scene gist together with more classic visual search tasks, offering a framework for considering how indi-

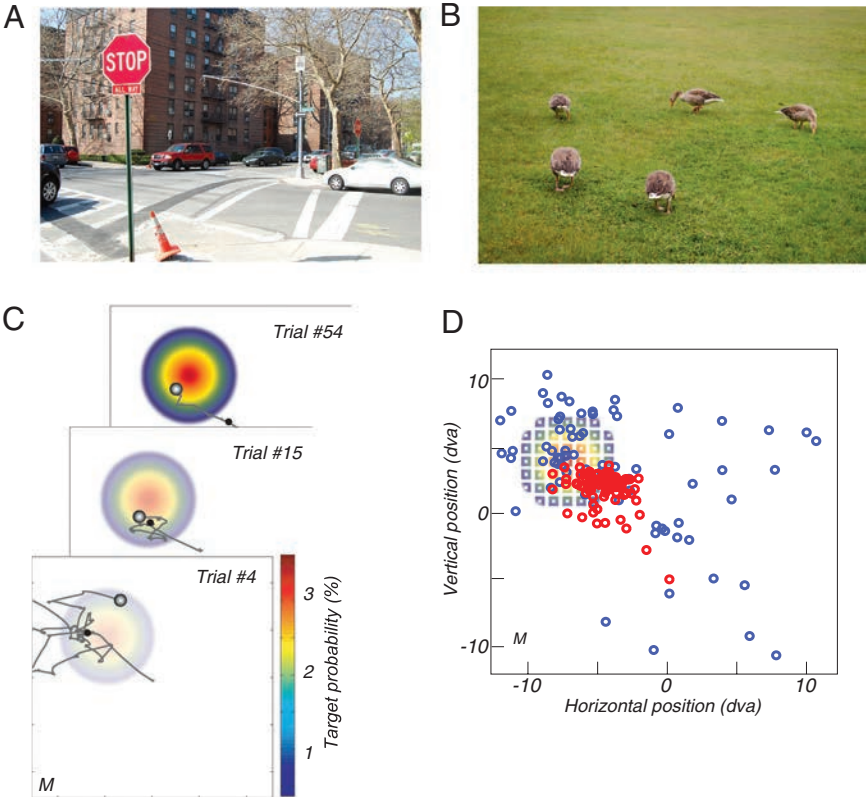


FIGURE 14.1 Visible and hidden search tasks. (A) An experienced pedestrian has prior knowledge of where to look for signs, cars, and sidewalks in this street scene. (B) Ducks foraging in a large expanse of grass. (C) A representation of the screen is superimposed with the hidden target distribution that is learned over the session as well as sample eye traces from three trials for participant M. The first fixation of each trial is marked with a black circle. The final and rewarded fixation is marked by a shaded grayscale circle. (D) The region of the screen sampled with fixation shrinks from the entire screen on early trials (blue circles; 87 fixations over the first five trials) to a region that approximates the size and position of the Gaussian-integer distributed target locations (squares, color proportional to the probability as given in A) on later trials (red circles; 85 fixations from trials 32–39). Fixation position data are from participant M. [Note: Figure can be viewed in color in the PDF version of this volume on the National Academies Press website, [www.nap.edu](http://www.nap.edu).]

viduals might use past experience to direct gaze in both real-world scenes as well as in the contrived scenarios of our laboratories.

Quite distinct from the literature on visual search is the literature on another sort of search that is commonly required of animals and people: foraging. Foraging agents seek food, which is often hidden in the environment in which they search (Fig. 14.1B). The search for hidden food rewards changes not only with the position of the reward, but also with the size of the distribution of rewards (Charnov, 1976). Other work has cast foraging behavior in terms of optimal search (Bénichou et al., 2005). What distinguishes foraging from visual search tasks is that visual search tasks have visible cues that drive search, in addition to contextual information that specifies probable target location. To make visual search more like foraging, we can strip the visible cues from visual search. A visual search task devoid of visual cues would allow us to determine whether there are underlying commonalities between these two types of search and whether general principles of search might emerge from such an investigation.

The importance of searching for hidden and even invisible targets is underscored by human participants engaged in large-scale exploration approximating animal foraging (Gilchrist et al., 2001; Smith et al., 2005). In one such paradigm (Smith et al., 2005), children were told to explore a room with a floor composed of box-like floor tiles, one of which contained a reward item. Interestingly, children explored the environment differently when they were instructed to search with their nondominant hand than with their dominant hand. Specifically, more “revisits” were necessary in the nondominant hand condition. This result suggests that learning and motor effort factor into performance on tasks that might seem to be automatic, which suggests methods for modeling foraging-like behavior. The additional motor effort that would be required to reduce metabolically expensive revisits in a foraging scenario seemed to have engaged memory systems to a greater degree than what is typically observed in traditional “visual” search tasks.

The reinforcement-learning (RL) framework has become widely accepted for modeling performance in tasks involving a series of movements leading to reward (Sutton, 1988; Montague and Sejnowski, 1994). In addition, for organisms across many levels of complexity, RL has been shown to be an appropriate framework to consider adaptive behavior in complex and changing environments (Niv et al., 2002; Lee et al., 2012). Here we describe performance in our task in terms of an RL perspective. Participants’ learning trajectories were well characterized by a simple RL model that maintained and continually updated a reward map of locations. The RL model made further predictions concerning sensitivity to recent experience that were confirmed by the data. The asymptotic performance of both the participants and the RL model approached optimal

performance characterized by an ideal-observer theory assuming perfect knowledge of the static target distribution and independently chosen fixations. These two complementary levels of explanation show how experience in a novel environment drives visual search in humans.

## RESULTS

### Humans Rapidly Learn to Find Hidden Targets

In visual search, previous experiments failed to isolate completely the visual appearance of a target from the learned location of the reward; in all cases a visual indication of a target, or a memory of a moments-ago visible target (Stritzke and Trommershäuser, 2007) and its surroundings, were available to guide the movement. To understand how participants learn where to look in a novel scene or context where no relationship exists between visual targets and associated rewards or penalties, we designed a search task in which participants were rewarded for finding a hidden target, similar to the scenario encountered by a foraging animal (Fig. 14.1C).

Participants repeatedly searched a single unfamiliar scene (context) for a target. However, to study the role of task knowledge in guiding search apart from the visual cues ordinarily used to identify a target, the target was rendered invisible. The participants' task was to explore the screen with their gaze and find a hidden target location that would sound a reward tone when fixated. Unbeknownst to each participant, the hidden target position varied from trial to trial and was drawn from a Gaussian distribution with a centroid and spread (target mean and SD, respectively) that was held constant throughout a session (Fig. 14.1C).

At the start of a session, participants had no prior knowledge to inform their search; their initial search was effectively "blind." As the session proceeded participants accumulated information from gaining reward or not at fixation points and improved their success rate by developing an expectation for the distribution of hidden targets and using it to guide future search (Fig. 14.1D).

After remarkably few trials, participants gathered enough information about the target distribution to direct gaze efficiently near the actual target distribution, as illustrated by one participant's data in Fig. 14.1C and D. We observed a similar pattern of learning for all participants: Early fixations were broadly scattered throughout the search screen; after approximately a dozen trials, fixations narrowed to the region with high target probability.

A characterization of this effect for all participants is shown in Fig. 14.2A. The average distance from the centroid of the target distribution to individual fixations in a trial drops precipitously over roughly the

first dozen trials. Fig. 14.2A shows this distance for all participants in the 2° target spread condition. The asymptotic distance from centroid increased monotonically with the target spread (Table 14.1).

A measure of search spread is the SD of the set of fixations in a trial. The search spread was initially broad and narrowed as the session progressed, as shown in Fig. 14.2B for all participants in the 2° target-spread condition. The asymptotic search spread monotonically increased with the target-spread condition (Table 14.1). These data suggest that participants estimated the spread of the hidden target distribution and adjusted their search spread accordingly. Also, the median number of fixations that participants made to find the target (on target-found trials) decreased rapidly within a session to reach an asymptote (Fig. 14.2C).

### Humans Approach Ideal-Observer Performance

We now consider the behavior of participants once performance had stabilized. Taking trials 31–60 to reflect asymptotic behavior, we examined the efficiency of human search in comparison with a theoretical optimum. An ideal observer was derived for the Hidden Target Search Task assuming that fixations are independent of one another and that the target distribution is known, and the expected number of trials is minimized. The dashed lines in Fig. 14.2 mark ideal-observer performance. Ideal search performance requires a distribution of planned fixation “guesses” that is  $\sqrt{2}$  broader than the target distribution itself (Snider, 2011). As seen in Fig. 14.2B and C, the performance of participants hovered around this ideal search distribution after about a dozen trials.

Subjects showed a  $\sim 1^\circ$  bias toward the center of the screen relative to the target distribution, but the calculation of the ideal behavior assumed subjects searched symmetrically around the center of the target distribution. Although the addition of the bias makes the math untenable analytically, a simulated searcher approximated the expected number of saccades required to find a target with a systematic  $1^\circ$  bias (Fig. 14.3). There was essentially no change in the predicted number of saccades or the search spread (location of the minimum in Fig. 14.3), except for the case of the  $0.75^\circ$  target distribution, where the optimum shifted from a search spread of  $0.56^\circ$  to  $0.85^\circ$ . Intuitively, the effect of bias was small because the bias was less than the  $2^\circ$  target radius. Nonetheless, at a 95 percent confidence level across the three target distributions, the number of steps, search spread, and step size all qualitatively and quantitatively match the predictions assuming the number of saccades was minimized.

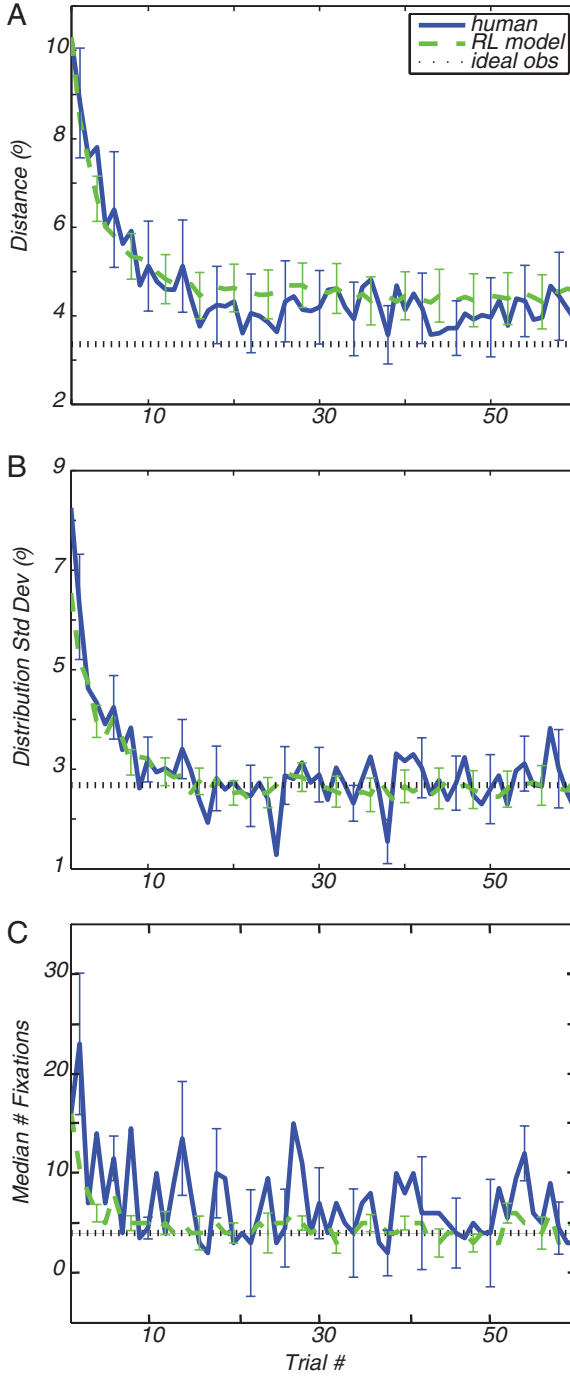




TABLE 14.1 Performance at Asymptote of Learning for Participants, the Ideal-Observer Theory, and a Reinforcement-Learning Model

Target Spread Condition, deg	Mean Distance from Target Centroid to Fixations on Trials 31–60, deg	Search Spread on Trials 31–60, deg
Participant data		
0.75	1.97	1.14
2.00	4.08	2.80
2.75	4.39	3.70
Ideal-observer theory		
0.75	0.70	0.56
2.00	3.36	2.68
2.75	4.74	3.78
Reinforcement-learning model		
0.75	3.21	1.56
2.00	4.46	2.61
2.75	6.07	4.29

NOTE: Data, theory, and model statistics for the mean fixation distance and search spread for 0.75-, 2.0-, and 2.75-degree target distribution conditions.

FIGURE 14.2 Learning curves for hidden-target search task. (A) The distance between the mean of the fixation cluster for each trial to the target centroid, averaged across participants, is shown in blue and green and indicates the result of 200 simulations of the reinforcement-learning model for each participant’s parameters. The SEM is given for both. The ideal-observer prediction is indicated by the black dotted line. (B) The SD of the eye position distributions or “search spread” is shown for the average of all participants (blue) and the RL model (green) with SEM. The dashed line is the ideal-observer theoretical optimum in each case, assuming perfect knowledge of the target distribution. (C) The median number of fixations made to find the target on each trial is shown (blue) along with the RL model prediction (green) of fixation number. The SEM is shown for both. [Note: Figure can be viewed in color in the PDF version of this volume on the National Academies Press website, [www.nap.edu](http://www.nap.edu).]

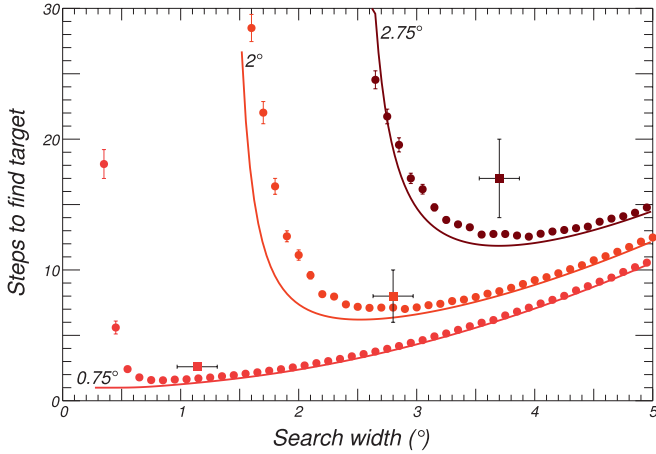


FIGURE 14.3 Optimal search model. Theoretical number of search steps to find the target for target distributions of size  $0.75^\circ$  (orange),  $2^\circ$  (red), and  $2.75^\circ$  (brown) was estimated by simulation (circles with mean and SEs from 100,000 trials per point) and from the theoretical calculation (solid lines). The simulation included the observed  $1^\circ$  bias seen in the subjects, but the theory lines did not. Solid boxes indicate the observed values for the subjects (mean and SE). With the added bias, the minimum moved slightly to the right but was only significant for the  $0.75^\circ$  target distribution. The cost in terms of extra saccades for nonoptimal search spreads (away from the minimum) was higher for the larger target distributions, and the comparatively shallow rise for search spreads above optimal meant that if subjects were to err, then they should tend toward larger spreads. Indeed, the tendency for larger spreads was evident as subjects started with large spreads and decreased toward the minimum (Fig. 14.2). The extra steps that subjects took to find the target for the  $2.75^\circ$  distribution (*Upper Right*) was consistent with the tendency toward small saccades even though they were quite close to the correct minimum: The largest saccades may have been broken up into multiple short saccades. [Note: Figure can be viewed in color in the PDF version of this volume on the National Academies Press website, [www.nap.edu](http://www.nap.edu).]

### Reinforcement-Learning Model Matches Human Learning

In addition to the ideal-observer theory, which characterizes the asymptotic efficiency of human search, we developed a complementary, mechanistic account that captured the learning, individual differences, and dynamics of human behavior. RL theory, motivated by animal learning and behavioral experiments (Yu and Cohen, 2008), suggests a simple and intuitive model that constructs a value function mapping locations in space to expected reward. The value function is updated after each fixa-

tion based on whether or not the target is found and is used for selecting saccade destinations that are likely to be rewarded.

We augmented this intuitive model with two additional assumptions: First, each time a saccade is made to a location, the feedback obtained generalized to nearby spatial locations; second, we incorporated a proximity bias that favored shorter saccades. A preference for shorter saccades was present in the data and has been noted by other researchers (Yarbus, 1967; Sutton and Barto, 1998), some of whom have shown that it can override knowledge that participants have about the expected location of a target (Rayner, 1998). Incorporating a proximity bias into the model changed the nature of the task because the choice of the next fixation became dependent on the current fixation. Consequently, participants must plan fixation sequences instead of choosing independent fixations.

We modeled the task using temporal difference methods (Sutton, 1988), which are particularly appropriate for Markovian tasks in which sequences of actions lead to reward (*Reinforcement Learning Model* gives details). The model's free parameters were fitted to each subject's sequence of fixations for each of the first 20 trials. Given these parameters, the model was then run in generative mode from a de novo state to simulate the subject performing the task.

Fig. 14.2 shows the mean performance of the model side by side with the mean human performance. The model also predicted an asymptotic search spread that increased with the target spread (Table 14.1), consistent with the participants' aggregate performance. Similar to the human performance observed in Fig. 14.2A, the RL model approaches, but does not reach, the theoretical asymptote. Like the human participants, the RL model is responsive to nonstationarity in the distribution, whereas the ideal-observer theory assumes that the distribution is static. In addition, the model accounted for individual differences (*Reinforcement Learning Model*). There are several reasons why the observed consistency between participants and simulations may be more than an existence proof and could provide insight into the biological mechanisms of learning (Araujo et al., 2001). The RL model itself had emergent dynamics that were reflected in the human behavior (Fig. 14.4 and sequential effects discussed below). Also the criterion used to train the model was the likelihood of a specific fixation sequence. A wide range of statistical measures quite distinct from the training criterion were used to compare human and model performance: mean distance from target centroid, SD of the distribution of eye movements, and the median number of fixations (Fig. 14.2). Finally, only the first 20 trials were used to train the model, but all of the comparisons shown in Table 14.1 were obtained from trials 31–60.

Fig. 14.2 suggests that participants acquire the target distribution in roughly a dozen trials and then their performance is static. However, in the

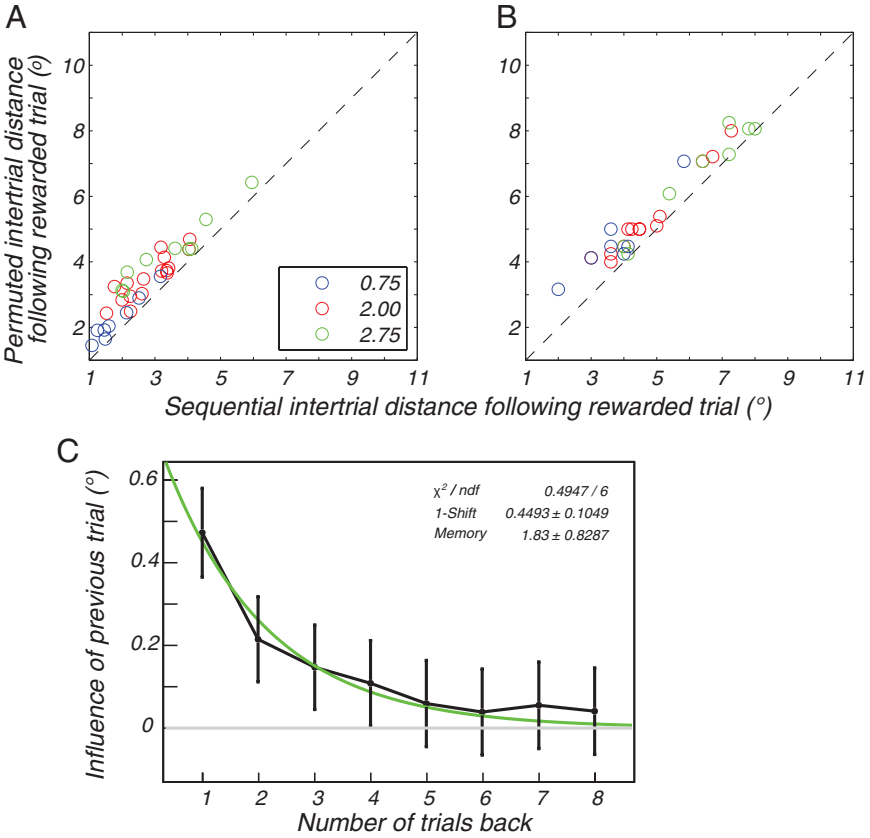


FIGURE 14.4 Sequential effects in the human data and predictions of the RL model. (A) For each subject, we plot the mean sequential intertrial distance (the distance between the final fixation on trial  $n$  and the first fixation on trial  $n + 1$  when trial  $n$  yields a reward) versus the permuted intertrial distance (the distance between the final fixation on a trial and the first fixation of another randomly drawn trial). Each circle denotes a subject, and the circle color indicates the target-spread condition (blue,  $\sigma = 0.75$ ; red,  $\sigma = 2.00$ ; green,  $\sigma = 2.75$ ). Consistent with the model prediction (B), the sequential intertrial distance is reliably shorter than permuted intertrial distance, as indicated by the points lying above the diagonal. All intertrial distances are larger in the model, reflecting a greater degree of exploration than in the participants, but this mismatch is orthogonal to the sequential effects. (C) The effect of previous trials on search in the current trial is plotted as a function of the number of trials back. An exponential fit to the data is shown in green. [Note: Figure can be viewed in color in the PDF version of this volume on the National Academies Press website, [www.nap.edu](http://www.nap.edu).]

RL model the value function is adjusted after each fixation, unabated over time. A signature of this ongoing adjustment is a sequential dependency across trials—specifically, a dependency between one trial’s final fixation and the next trial’s initial fixation. Dependencies were indeed observed in the data throughout a session (Fig. 14.4A), as predicted by the model (Fig. 14.4B) and explained some of the trial-to-trial variability in performance (Fig. 14.2 and *Reinforcement Learning Model*). Participants were biased to start the next trial’s search near found target locations from recent trials. The influence of previous trials decreases exponentially; the previous two, or possibly three, trials influenced the current trial’s saccade choice (Fig. 14.4C). This exponential damping of previous trials’ influence is approximated by the memoryless case (Snider, 2011), allowing both the RL model and ideal planner to coexist asymptotically.

### Bimodal Distribution of Saccade Lengths

Our motivation in designing the hidden target search task was to link the visual search and foraging literatures. Performance in our task had features analogous to those found in the larger context of animal foraging (Fig. 14.5). Although individual trials look like Lévy flights—a mixture of fixation and sporadic large excursions that are known to be optimal in some cases of foraging behavior (Schultz et al., 1997; Humphries et al., 2010; James et al., 2011)—the length distribution of all straight-line segments is not Lévy-like, but separates into two distinct length scales like the intermittent search popularized by Bénichou et al. (2005). The shorter length scale, fixations less than about  $1^\circ$ , corresponds to a local power-law search with a very steep exponent, making it a classic random walk that densely samples the local space. That local search is combined with the larger, but rarer, saccades represented by the peaked hump at step sizes larger than  $1^\circ$ . These are the distinct choices from the planned distribution described already (i.e., the guess distribution or value function). The distinctive knee shape in Fig. 14.5 is similar to that found in other demanding visual search tasks (Snider, 2011), as well as intermittent foraging by a wide range of animals (Bénichou et al., 2005; Humphries et al., 2010).

## DISCUSSION

Human search performance can be put into the more general context of animal foraging, which has close connections with RL models (Niv et al., 2002) and optimal search theory (Charnov, 1976). The hidden target search task introduced here has allowed us to separate the influence of external cues from internal prior information for seeking rewards in a novel environment (Viswanathan et al., 1999). In our hidden target search

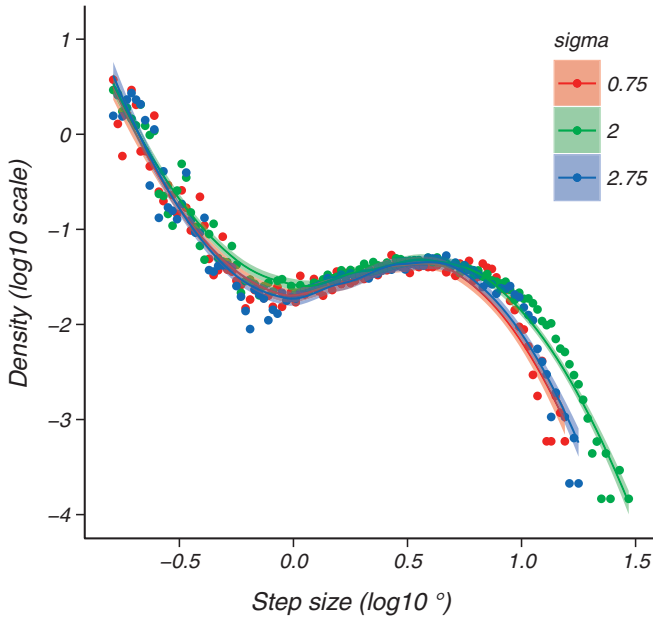


FIGURE 14.5 Length distributions of saccades in the hidden target task. A turning-point algorithm applied to raw eye movement data yields a distribution of step sizes for all participants (*Reinforcement Learning Model* gives details). Very small “fixational” eye movements comprise the left side of the plot and large larger saccadic jumps on the right for three different sizes of target distribution. The points and lines (Loess fits with 95 percent confidence interval shading) for each search distribution size, all share a similar shape, particularly a bend at step sizes approaching  $1^\circ$  of visual angle. [Note: Figure can be viewed in color in the PDF version of this volume on the National Academies Press website, [www.nap.edu](http://www.nap.edu).]

task, participants explored a novel environment and quickly learned to align their fixations with the region of space over which invisible targets were probabilistically distributed. After about a dozen trials, the fixation statistics came close to matching those obtained by an ideal-observer theory. This near-match allowed us to cast human performance as optimal memory-free search with perfect knowledge of the target distribution. As a complement to the ideal-observer theory that addresses asymptotic performance, we developed a mechanistic account of trial-to-trial learning from reinforcement. Our RL model characterized the time course of learning, attained an asymptote near ideal-observer performance, and tied the problem of visual search to a broader theory of motivated learning.

## Natural Environments

The ideal-observer and reinforcement-learning frameworks provide the foundation for a broader theoretical perspective on saccade choice during natural vision, in which people learn to search in varied contexts for visible targets, where visual features of the scene are clearly essential. In a Bayesian framework, the subjects in our task learned the prior distribution of the hidden targets. In a natural environment, the prior distribution would be combined with visual information to determine the posterior distribution, from which saccadic targets are generated.

Naturalistic environments are nonstationary. For example, an animal foraging for food may exhaust the supply in one neighborhood and have to move on to another. A searcher must be sensitive to such changes in the environment. Sequential dependencies (Fig. 14.4) are a signature of this sensitivity (Fecteau and Munoz, 2003; Wilder et al., 2011; Adams et al., 2012): Recent targets influence subsequent behavior, even after the searcher has seemingly learned the target distribution, as reflected in asymptotic performance. Sequential dependencies were predicted by the RL model, which generated behavior remarkably close to that of the participants as a group, and also captured individual idiosyncrasies (*Reinforcement Learning Model*). Sensitivity to nonstationary environments can explain why our participants and the RL model attained an asymptotic search distribution somewhat further from the target centroid than is predicted by an ideal-observer theory premised on stationarity.

One of the most impressive feats of animal foraging is matching behavior. Herrnstein's matching law (Herrnstein, 1961) describes how foraging animals tend to respond in proportion to the expected value of different patches. Matching behavior has been studied in multiple species from honey bees to humans (Bradshaw et al., 1976; Greggers and Mauelshagen, 1997; Lau and Glimcher, 2005; Gallistel et al., 2007). However, many of these laboratory studies effectively remove the spatial element of foraging from the task by looking at different intervals of reinforcement on two levers or buttons; in this setting, animals quickly detect changes in reinforcement intervals (Mark and Gallistel, 1994) and the motor effort in switching between spatial patches has been examined (Baum, 1982). In nature, foraging is spatially extended, and the hidden-target search paradigm could serve as an effective environment for examining an explicitly spatial foraging task in the context of matching behavior. For example, a version of our hidden-target search paradigm with a bimodal distribution could explore changeover behavior and motor effort by varying the sizes of the two distributions and distance between them (Baum, 1982).

## Neural Basis of Search

The neurobiology of eye movement behavior offers an alternative perspective on the similarities of visual search behavior and foraging. The question of where to look next has been explored neurophysiologically, and cells in several regions of the macaque brain seem to carry signatures of task components required for successful visual search. The lateral intraparietal (LIP) area and the superior colliculus (SC) are two brain regions that contain a priority map representing locations of relevant stimuli that could serve as the target of the next saccade. Recordings in macaque area LIP and the SC have shown that this priority map integrates information from both external (“bottom-up”) and internal (“top-down”) signals in visual search tasks (Fecteau and Munoz, 2006; Bisley and Goldberg, 2010).

Recently, Bisley and coworkers (Mirpour et al., 2009) have used a foraging-like visual search task to show that area LIP cells differentiated between targets and distracters and kept a running estimate of likely saccade goal payoffs. Area LIP neurons integrate information from different foraging-relevant modalities to encode the value associated with a movement to a particular target (Platt et al., 1999; Klein et al., 2008). The neural mechanisms serving patch stay-leave foraging decisions have recently been characterized in a simplified visual choice task (Hayden et al., 2011), providing a scheme for investigations of precisely how prior information and other task demands mix with visual information available in the scene. Subthreshold microstimulation in area LIP (Mirpour et al., 2010) or the SC (Carello and Krauzlis, 2004) also biases the selection saccades toward the target in the stimulated field. Taken together, these results suggest that area LIP and the SC might be neural substrates mediating the map of likely next saccade locations in our task, akin to the value map in our RL model.

We asked how subjects learn to choose valuable targets in a novel environment. Recent neurophysiological experiments in the basal ganglia provide some suggestions on how prior information is encoded for use in choosing the most valuable saccade target in a complex environment (Nakahara and Hikosaka, 2012). Hikosaka and coworkers (Yasuda et al., 2012) have identified signals related to recently learned, and still labile, value information for saccade targets in the head of the caudate nucleus and more stable value information in the tail of the caudate and substantia nigra, pars reticulata. Because the cells carrying this stable value information seem to project preferentially to the SC, these signals are well placed to influence saccade choices through a fast and evolutionarily conserved circuit for controlling orienting behavior. These results provide a neurophysiological basis for understanding how experience is learned and consolidated in the service of the saccades we make to gather information about our environment about three times each second.



## CONCLUSIONS

In our eye-movement search task, subjects learned to choose saccade goals based on prior experience of reward that is divorced from specific visual features in a novel scene. The resulting search performance was well described by an RL model similar to that used previously to examine both foraging animal behavior and neuronal firing of dopaminergic cells. In addition, the search performance approached the theoretical optimum for performance on this task. By characterizing how prior experience guides eye movement choice in novel contexts and integrating it with both model and theory, we have created a framework for considering how prior experience guides saccade choice during natural vision. The primate oculomotor system has been well studied, which will make it possible to uncover the neural mechanisms underlying the learning and performance of the hidden-target task, which may be shared with other search behaviors.

## METHODS

We defined a spatial region of an image as salient by associating it with reward to examine how participants used their prior experience of finding targets to direct future saccades. We took advantage of the fact that the goal of saccadic eye movements is to obtain information about the world and asked human participants to “conduct an eye movement search to find a rewarded target location as quickly as possible.” Participants were also told that they would learn more about the rewarded targets as the session progressed and that they should try to find the rewarded target location as quickly as possible. The rewarded targets had no visual representation on the screen and were thus invisible to the subject. The display screen was the same on each trial within a session and provided no information about the target location. The location and the spread of the rewarded target distribution were varied with each session.

Each trial began with a central fixation cross on a neutral gray screen with mean luminance of  $36.1 \text{ cd/m}^2$  (Fig. 14.1). The search screen spanned the central  $25.6^\circ$  of the subject’s view while seated with his or her head immobilized by a bite bar.

Participants initiated each trial with a button press indicating that they were fixating the central cross. The same neutral gray screen served as the search screen after 300 ms of fixation of the cross. Once the fixation cross disappeared, participants had 20 seconds to find the rewarded location for that trial before the fixation screen returned. On each trial an invisible target was drawn from a predefined distribution of possible targets. The shape of the distribution was Gaussian with the center at an integer number of degrees from the fixation region (usually  $\pm 6^\circ$  in  $x$  and

$y$ ) and spread held fixed over each experimental session. The targets only occurred at integer values of the Gaussian. The probability associated with a rewarded target location varied between 4 percent and 0.1 percent and was given by the spread of the distribution ( $0.75^\circ$ ,  $2^\circ$ , and  $2.75^\circ$  SD). When a subject's gaze landed within  $2^\circ$  of the target in both the  $x$  and  $y$  directions, a reward tone marked the successful end of the trial. For the target to be "found," fixation (monitored in real time as detailed below) needed to remain steady within the target window for at least 50 ms. This duration ensured that the target was never found simply by sweeping through during a saccade. If at the end of 20 seconds the target was not found, the trial ended with no tone and a fixation cross appeared indicating the beginning of a new trial.

Trial timing and data collection were managed by the TEMPO software system (Reflective Computing) and interfaced with the stimulus display using an extension of the Psychophysics Toolbox (Brainard, 1997) running under MATLAB (MathWorks). Eye movement data were obtained using a video-based eye tracker (ISCAN), sampled at 240 Hz for humans. Eye data were calibrated by having the participants look at stimuli at known locations. Eye movements were analyzed offline in MATLAB. We detected saccades and blinks by using a conservative velocity threshold ( $40^\circ/\text{s}$  with a 5-ms shoulder after each saccade) after differentiating the eye position signals. Periods of steady fixation during each trial were then marked and extracted for further analyses. Eye positions off of the search screen were discounted from analysis. Visual inspection of individual trials confirmed that the marked periods of fixation were indeed free from saccades or blinks.

### Turning Points

In addition to saccades identified by speed criteria, the eye tracking data were processed to estimate the step-size distribution of all eye movements, even within a fixation. To that end, blinks were first removed by removing samples off the screen. Next, we considered the data points three at a time,  $\mathbf{x}_{t-1}$ ,  $\mathbf{x}_t$ , and  $\mathbf{x}_{t+1}$ , where  $\mathbf{x}$  are the 2D data points and  $t$  indexes the time samples, to construct two segments of the eye track  $\mathbf{a} = \mathbf{x}_{t-1} - \mathbf{x}_t$  and  $\mathbf{b} = \mathbf{x}_t - \mathbf{x}_{t+1}$ . We then tested whether the cosine of the angle between these two was greater or less than 0.95. If the cosine was greater than 0.95, then the center point,  $\mathbf{x}_t$ , was marked as a "turning" point. In addition, some of the large steps slowly curved and this introduced extraneous points (i.e., dividing a long step into two short steps). To overcome this problem, we took advantage of the fact that two long steps almost never occur one after the other without a dense fixation region in between, and any point with no neighbors within  $0.5^\circ$  was assumed to be extraneous

and was removed. This resulted in points at which the eye made a significant deviation from ballistic motion and was used to generate the step size distributions in Fig. 14.5.

### **ACKNOWLEDGMENTS**

We thank Krista Kornylo and Natalie Dill for technical assistance. This work was supported in part by National Science Foundation (NSF) Grant SBE 0542013 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center grant (to L.C., J.S., M.C.M., and T.J.S.), a Blasker Rose-Miah grant from the San Diego Foundation (to L.C.), Office of Naval Research Multidisciplinary University Research Initiative Award N00014-10-1-0072 (to J.S. and T.J.S.), and Howard Hughes Medical Institute (T.J.S.).