Learning where to look for a hidden target

Leanne Chukoskie[a], Joseph Snider[a], Michael C. Mozer[a,b,c], Richard J. Krauzlis[d], and Terrence J. Sejnowski[a,e,f,g,1]

[a]Institute for Neural Computation and [g]Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093; [b]Department of Computer Science and [c]Institute of Cognitive Science, University of Colorado, Boulder, CO 80309; [d]Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD 20892; and [e]Howard Hughes Medical Institute and [f]Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037

Edited by Francisco J. Ayala, University of California, Irvine, CA, and approved May 1, 2013 (received for review February 16, 2013)

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 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. 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.

ideal observer | oculomotor | reinforcement learning | saccades

The 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 (1–9). This ability to use the "gist" of an image (3, 4) 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 (10) and it becomes clear that despite the large body of research on how image salience guides gaze (2, 11), learned spatial associations are perhaps just as important for effectively engaging our visual environment (10, 12, 13). 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 (14). In fact, finding something you seek may be intrinsically rewarding (15). 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 (16, 17). 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 (18), are different (19) and perhaps also differently valuable (20) 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 (21) and Yarbus (22) 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 (23) have replicated this result for the different tasks of visual search and image memorization. However, Wolfe and coworkers (24), 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 (4). For example, when shown a street scene, an observer would immediately know where to look for street signs, cars, and pedestrians (Fig. 1*A*).

Castelhano and Heaven (9) 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 (25–27) to hone the aim of a saccadic

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "In the Light of Evolution VII: The Human Mental Machinery," held January 10–12, 2013, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and audio files of most presentations are available on the NAS Web site at www.nasonline.org/evolution_vii.

Author contributions: L.C., J.S., M.C.M., R.J.K., and T.J.S. designed research; L.C., J.S., and M.C.M. performed research; L.C., J.S., M.C.M., R.J.K., and T.J.S. analyzed data; and L.C., J.S., M.C.M., R.J.K., and T.J.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: terry@salk.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1301216110/-/DCSupplemental.

eye movement. A recent review and commentary from Wolfe et al. (28) 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 individuals 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. 1*B*). 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 (29). Other work has cast foraging behavior in terms of optimal search (30). 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 (31, 32). In one such paradigm (32), 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 (33, 34). 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 (35, 36). Here we describe performance in our task in terms of a 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.

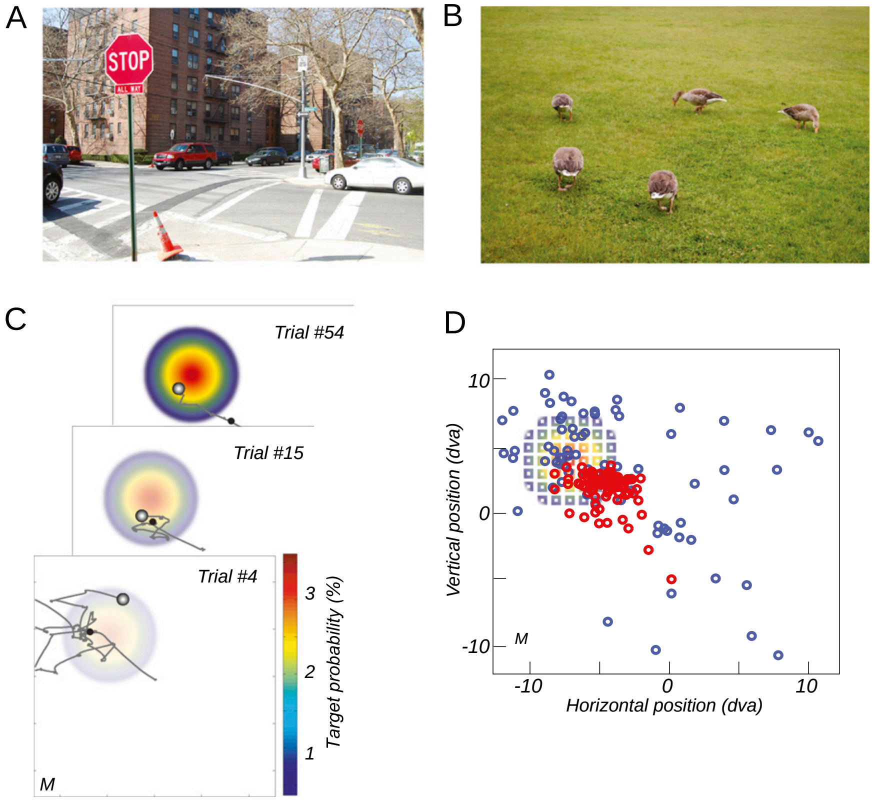

Fig. 1. Visible and hidden search tasks. (*A*) An experienced pedestrian has prior knowledge of where to look for signs, cars, and sidewalks in this street scene. (*B*) Ducks foraging in a large expanse of grass. (*C*) A representation of the screen is superimposed with the hidden target distribution 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.

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 (26) 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. 1*C*).

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. 1*C*).

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. 1*D*).

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. 1 *C* 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. 2*A*. 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. 2*A* shows this distance for all participants in the 2° target spread condition. The asymptotic distance from centroid increased monotonically with the target spread (Table 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. 2*B* for all participants in the 2° target-spread condition. The asymptotic search spread monotonically increased with the target-spread condition (Table 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. 2*C*).

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 (Fig. S1 and Table S1). The dashed lines in Fig. 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 (37). As seen in Fig. 2 *B* and *C*, the performance of participants hovered around this ideal search distribution after about a dozen trials.

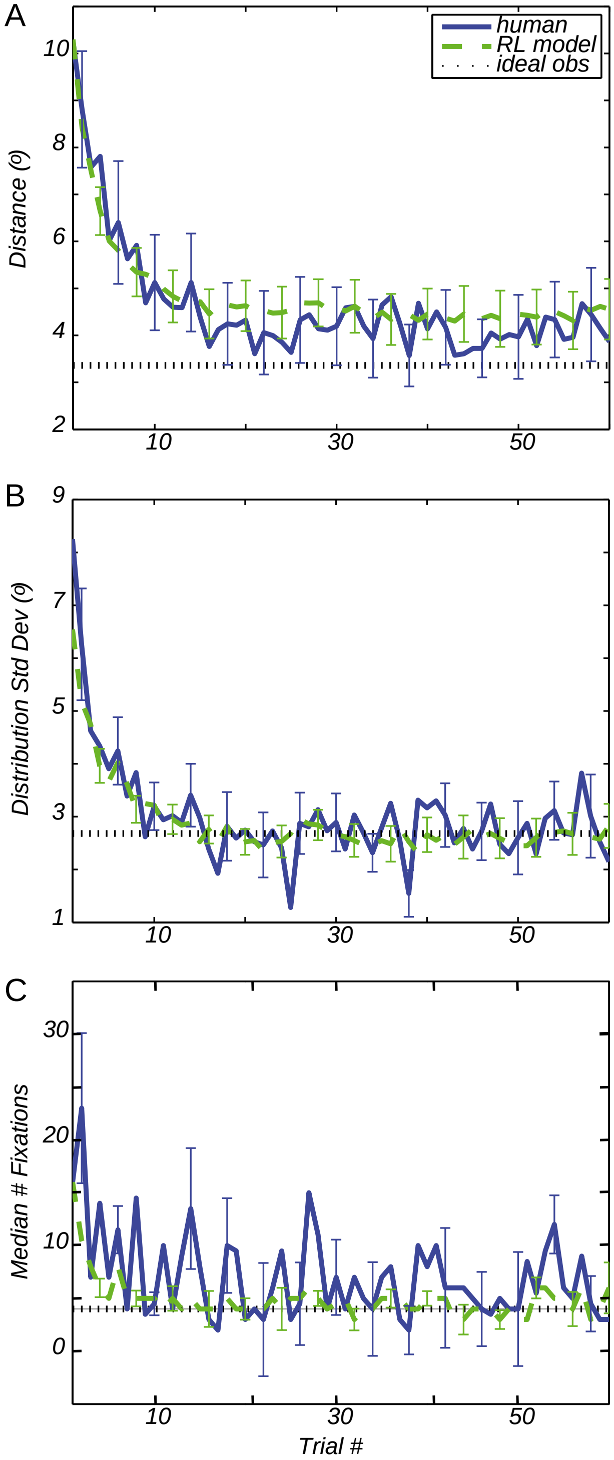

Fig. 2. Learning curves for hidden-target search task. (*A*) The distance between the mean of the fixation cluster for each trial to the target centroid, averaged across participants, is shown in blue and green and indicates the result of 200 simulations of the reinforcement-learning model for each participant's 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.

Table 1. Performance at asymptote of learning for participants, the ideal-observer theory, and a reinforcement-learning model

Target spread condition, °	Mean distance from target centroid to fixations on trials 31–60, °	Search spread on trials 31–60, °
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

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

Subjects showed a ~1° bias toward the center of the screen relative to the target distribution (Table S2), 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° bias (Fig. 3). There was essentially no change in the predicted number of saccades or the search spread (location of the minimum in Fig. 3), except for the case of the 0.75° target distribution, where the optimum shifted from a search spread of 0.56° to 0.85°. Intuitively, the effect of bias was small because the bias was less than the 2° target radius. Nonetheless, at a 95% 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.

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 (38), 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 fixation 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 (Fig. S2) and has been noted by other researchers (22, 39), some of whom have shown that it can override knowledge that participants have about the expected location of a target (40). 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 (33), which are particularly appropriate for Markovian tasks in which sequences of actions lead to reward (*Reinforcement Learning Model* and Figs. S2 and S3 give details). The model's free parameters were fit 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. 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 1), consistent with the participants' aggregate performance. Similar to the human performance observed in Fig. 2*A*, 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

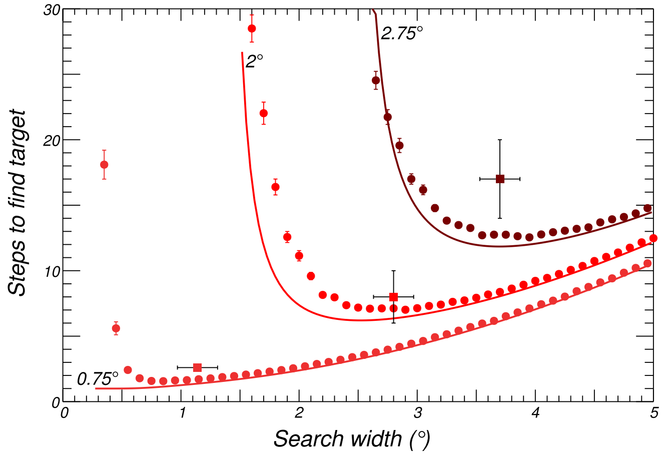

Fig. 3. Optimal search model. Theoretical number of search steps to find the target for target distributions of size 0.75° (orange), 2° (red), and 2.75° (brown) was estimated by simulation (circles with mean and SEs from 100,000 trials per point) and from the theoretical calculation (solid lines) as detailed in Table S1 and *Supporting Information*. The simulation included the observed 1° 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° 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. 2). The extra steps that subjects took to find the target for the 2.75° distribution (*Upper Right*) was consistent with the tendency toward small saccades even though they were quite close to the correct minimum (Fig. S2): The largest saccades may have been broken up into multiple short saccades.

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 (41). The RL model itself had emergent dynamics that were reflected in the human behavior (Fig. 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. 2). Finally, only the first 20 trials were used to train the model, but all of the comparisons shown in Table 1 were obtained from trials 31–60.

Fig. 2 suggests that participants acquire the target distribution in roughly a dozen trials and then their performance is static. However, in the 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. 4*A*), as predicted by the model (Fig. 4*B*) and explained some of the trial-to-trial variability in performance (Fig. 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. 4*C*). This exponential damping of previous trials' influence is approximated by the memoryless case (37), 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. 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 (42–44)—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. (30). The shorter length scale, fixations less than about 1°, 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°. 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. 5 is similar to that found in other demanding visual search tasks (37), as well as intermittent foraging by a wide range of animals (30, 43).

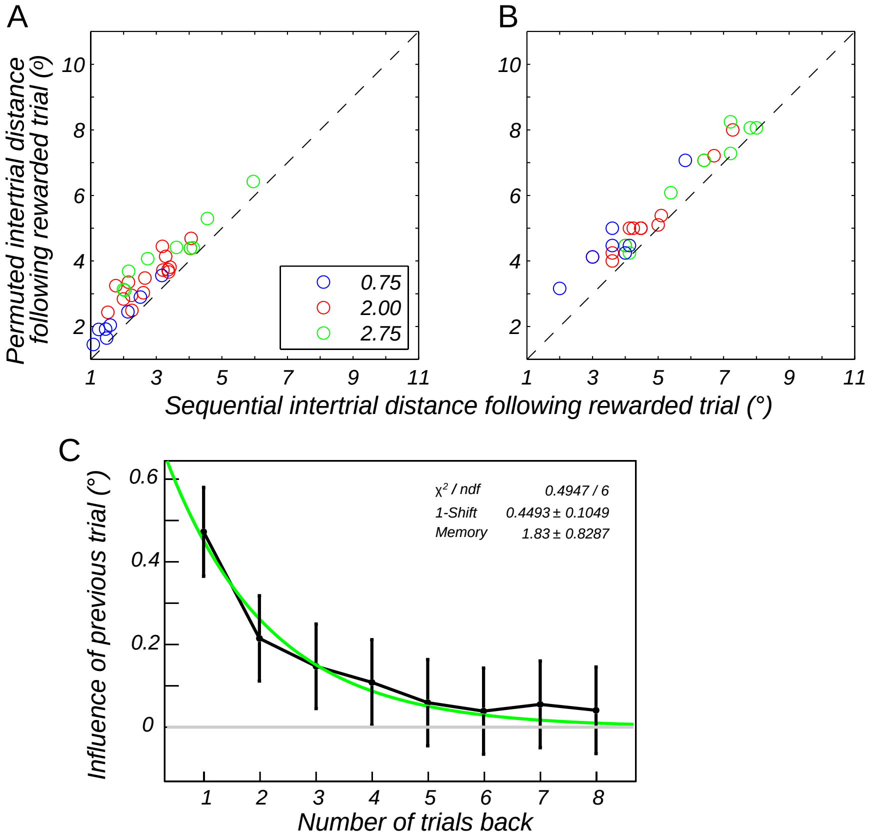

Fig. 4. Sequential effects in the human data and predictions of the RL model. (*A*) For each subject, we plot the mean sequential intertrial distance (the distance between the final fixation on trial n and the first fixation on trial $n + 1$ when trial n yields a reward) versus the permuted intertrial 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 are shown in green.

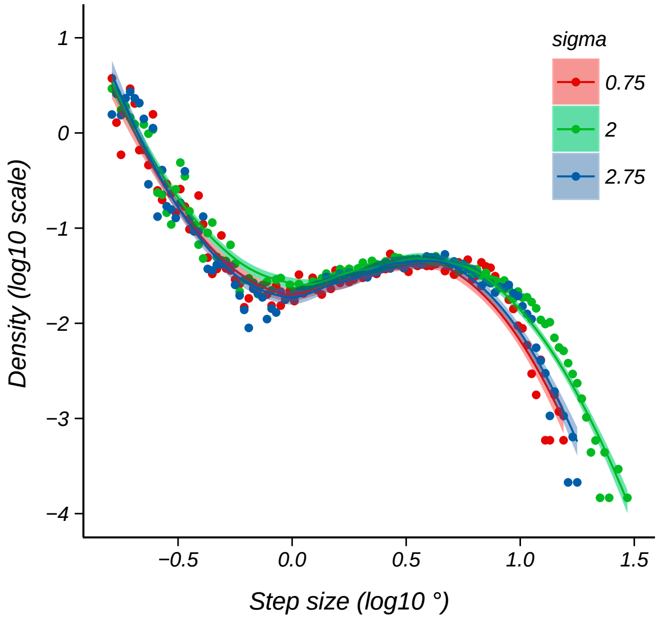

Fig. 5. Length distributions of saccades in the hidden target task. A turning point algorithm applied to raw eye movement data yields a distribution of step sizes for all participants (*Reinforcement Learning Model* gives details). Very small "fixational" eye movements comprise the left side of the plot and large larger saccadic jumps on the right for three different sizes of target distribution. The points and lines (Loess fits with 95% confidence interval shading) for each search distribution size, all share a similar shape, particularly a bend at step sizes approaching 1° of visual angle.

Discussion

Human search performance can be put into the more general context of animal foraging, which has close connections with RL models (36) and optimal search theory (29). 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 (45). In our hidden target search 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. 4) are a signature of this sensitivity (46–48): 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 (49) 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 (50–53). 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 (54) and the motor effort in switching between spatial patches has been examined (55). 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 (55).

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 interparietal (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 (56, 57).

Recently, Bisley and coworkers (58) 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 (59, 60). The neural mechanisms serving patch stay-leave foraging decisions have recently been characterized in a simplified visual choice task (61), 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 (62) or the SC (63) 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 (64). Hikosaka and coworkers (65) 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 cd/m^2 (Fig. 1). The search screen spanned the central 25.6° 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 s 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 $\pm 6°$ 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% and 0.1% and was given by the spread of the distribution (0.75°, 2°, and 2.75° SD). When a subject's gaze landed within 2° 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 s 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 (66) 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°/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° 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. 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.).

1. Potter MC (1975) Meaning in visual search. *Science* 187(4180):965–966.
2. Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res* 40(10-12):1489–1506.
3. Oliva A, Torralba A (2006) Building the gist of a scene: The role of global image features in recognition. *Prog Brain Res* 155:23–36.
4. Torralba A, Oliva A, Castelhano MS, Henderson JM (2006) Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychol Rev* 113(4):766–786.
5. Neider MB, Zelinsky GJ (2006) Scene context guides eye movements during visual search. *Vision Res* 46(5):614–621.
6. Rayner K, Castelhano MS, Yang J (2009) Eye movements when looking at unusual/weird scenes: Are there cultural differences? *J Exp Psychol Learn Mem Cogn* 35(1):254–259.
7. Võ ML, Henderson JM (2010) The time course of initial scene processing for eye movement guidance in natural scene search. *J Vis* 10(3):11–13.
8. Castelhano MS, Heaven C (2010) The relative contribution of scene context and target features to visual search in scenes. *Atten Percept Psychophys* 72(5):1283–1297.
9. Castelhano MS, Heaven C (2011) Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychon Bull Rev* 18(5):890–896.
10. Hayhoe M, Ballard D (2005) Eye movements in natural behavior. *Trends Cogn Sci* 9(4):188–194.
11. Parkhurst DJ, Niebur E (2003) Scene content selected by active vision. *Spat Vis* 16(2):125–154.
12. Tatler BW, Vincent BT (2009) The prominence of behavioural biases in eye guidance. *Vis Cogn* 17(6–7):1029–1054.
13. Chun MM, Jiang Y (1998) Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognit Psychol* 36(1):28–71.
14. Milstein DM, Dorris MC (2007) The influence of expected value on saccadic preparation. *J Neurosci* 27(18):4810–4818.
15. Xu-Wilson M, Zee DS, Shadmehr R (2009) The intrinsic value of visual information affects saccade velocities. *Exp Brain Res* 196(4):475–481.
16. Shadmehr R (2010) Control of movements and temporal discounting of reward. *Curr Opin Neurobiol* 20(6):726–730.
17. Shadmehr R, Orban de Xivry JJ, Xu-Wilson M, Shih TY (2010) Temporal discounting of reward and the cost of time in motor control. *J Neurosci* 30(31):10507–10516.
18. Reinagel P, Zador AM (1999) Natural scene statistics at the centre of gaze. *Network* 10(4):341–350.
19. Tatler BW, Baddeley RJ, Vincent BT (2006) The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Res* 46(12):1857–1862.
20. Açık A, Sarwary A, Schultze-Kraft R, Onat S, König P (2010) Developmental changes in natural viewing behavior: Bottom-up and top-down differences between children, young adults and older adults. *Front Psychol* 1:207.
21. Buswell GT (1935) *How People Look at Pictures: A Study of the Psychology of Perception in Art* (Univ of Chicago Press, Chicago).
22. Yarbus AL (1967) *Eye Movements and Vision* (Plenum, New York).
23. Castelhano MS, Mack ML, Henderson JM (2009) Viewing task influences eye movement control during active scene perception. *J Vis* 9(3):1–15.
24. Greene MR, Liu T, Wolfe JM (2012) Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Res* 62:1–8.

25. Schütz AC, Trommershäuser J, Gegenfurtner KR (2012) Dynamic integration of information about salience and value for saccadic eye movements. *Proc Natl Acad Sci USA* 109(19):7547–7552.
26. Stritzke M, Trommershäuser J (2007) Eye movements during rapid pointing under risk. *Vision Res* 47(15):2000–2009.
27. Geng JJ, Behrmann M (2005) Spatial probability as an attentional cue in visual search. *Percept Psychophys* 67(7):1252–1268.
28. Wolfe JM, Võ ML, Evans KK, Greene MR (2011) Visual search in scenes involves selective and nonselective pathways. *Trends Cogn Sci* 15(2):77–84.
29. Charnov EL (1976) Optimal foraging, the marginal value theorem. *Theor Popul Biol* 9(2):129–136.
30. Bénichou O, Coppey M, Moreau M, Suet P-H, Voituriez R (2005) Optimal search strategies for hidden targets. *Phys Rev Lett* 94(19):198101–198104.
31. Gilchrist ID, North A, Hood B (2001) Is visual search really like foraging? *Perception* 30(12):1459–1464.
32. Smith AD, Gilchrist ID, Hood BM (2005) Children's search behaviour in large-scale space: Developmental components of exploration. *Perception* 34(10):1221–1229.
33. Sutton RS (1988) Learning to predict by the method of temporal differences. *Mach Learn* 3(1):9–44.
34. Montague PR, Sejnowski TJ (1994) The predictive brain: Temporal coincidence and temporal order in synaptic learning mechanisms. *Learn Mem* 1(1):1–33.
35. Lee D, Seo H, Jung MW (2012) Neural basis of reinforcement learning and decision making. *Annu Rev Neurosci* 35:287–308.
36. Niv Y, Joel D, Meilijson I, Ruppin E (2002) Evolution of reinforcement learning in uncertain environments: A simple explanation for complex foraging beaviors. *Adapt Behav* 10(1):5–24.
37. Snider J (2011) Optimal random search for a single hidden target. *Phys Rev E Stat Nonlin Soft Matter Phys* 83(1 Pt 1):011105.
38. Yu AJ, Cohen JD (2008) Sequential effects: Superstition or rational behavior?, eds Koller D, Schuurmans D, Bengio Y, Bottou L *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA), Vol. 21, pp 1873–1880.
39. Sutton RS, Barto AG (1998) *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA), p xviii.
40. Rayner K (1998) Eye movements in reading and information processing: 20 years of research. *Psychol Bull* 124(3):372–422.
41. Araujo C, Kowler E, Pavel M (2001) Eye movements during visual search: The costs of choosing the optimal path. *Vision Res* 41(25–26):3613–3625.
42. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275(5306):1593–1599.
43. Humphries NE, et al. (2010) Environmental context explains Lévy and Brownian movement patterns of marine predators. *Nature* 465(7301):1066–1069.
44. James A, Plank MJ, Edwards AM (2011) Assessing Lévy walks as models of animal foraging. *J R Soc Interface* 8(62):1233–1247.
45. Viswanathan GM, et al. (1999) Optimizing the success of random searches. *Nature* 401 (6756):911–914.
46. Adams GK, Watson KK, Pearson J, Platt ML (2012) Neuroethology of decision-making. *Curr Opin Neurobiol* 22(6):982–989.
47. Fecteau JH, Munoz DP (2003) Exploring the consequences of the previous trial. *Nat Rev Neurosci* 4(6):435–443.
48. Wilder MH, Mozer MC, Wickens CD (2011) An integrative, experience-based theory of attentional control. *J Vis* 11(2), 10.1167/11.2.8.
49. Herrnstein RJ (1961) Relative and absolute strength of response as a function of frequency of reinforcement. *J Exp Anal Behav* 4:267–272.
50. Greggers U, Mauelshagen J (1997) Matching behavior of honeybees in a multiple-choice situation: The differential effect of environmental stimuli on the choice process. *Anim Learn Behav* 25(4):458–472.
51. Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. *J Exp Anal Behav* 84(3):555–579.
52. Bradshaw CM, Szabadi E, Bevan P (1976) Behavior of humans in variable-interval schedules of reinforcement. *J Exp Anal Behav* 26(2):135–141.
53. Gallistel CR, et al. (2007) Is matching innate? *J Exp Anal Behav* 87(2):161–199.
54. Mark TA, Gallistel CR (1994) Kinetics of matching. *J Exp Psychol Anim Behav Process* 20(1):79–95.
55. Baum WM (1982) Choice, changeover, and travel. *J Exp Anal Behav* 38(1):35–49.
56. Bisley JW, Goldberg ME (2010) Attention, intention, and priority in the parietal lobe. *Annu Rev Neurosci* 33:1–21.
57. Fecteau JH, Munoz DP (2006) Salience, relevance, and firing: A priority map for target selection. *Trends Cogn Sci* 10(8):382–390.
58. Mirpour K, Arcizet F, Ong WS, Bisley JW (2009) Been there, seen that: A neural mechanism for performing efficient visual search. *J Neurophysiol* 102(6):3481–3491.
59. Klein JT, Deaner RO, Platt ML (2008) Neural correlates of social target value in macaque parietal cortex. *Curr Biol* 18(6):419–424.
60. Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400(6741):233–238.
61. Hayden BY, Pearson JM, Platt ML (2011) Neuronal basis of sequential foraging decisions in a patchy environment. *Nat Neurosci* 14(7):933–939.
62. Mirpour K, Ong WS, Bisley JW (2010) Microstimulation of posterior parietal cortex biases the selection of eye movement goals during search. *J Neurophysiol* 104(6): 3021–3028.
63. Carello CD, Krauzlis RJ (2004) Manipulating intent: Evidence for a causal role of the superior colliculus in target selection. *Neuron* 43(4):575–583.
64. Nakahara H, Hikosaka O (2012) Learning to represent reward structure: A key to adapting to complex environments. *Neurosci Res* 74(3-4):177–183.
65. Yasuda M, Yamamoto S, Hikosaka O (2012) Robust representation of stable object values in the oculomotor Basal Ganglia. *J Neurosci* 32(47):16917–16932.
66. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.

Supporting Information

Chukoskie et al. 10.1073/pnas.1301216110

Ideal Observer Model

The general problem to solve is given a target distribution $T(\vec{x})$ describing the probability a target is at location $\vec{x}$ and a search window with size R, estimate the guess distribution $G(\vec{x})$ from which random points are sampled until the target is within R of the random point. This has been solved previously (1), and we will just reproduce a simplified version here.

Along with existence of the guess and target distributions, $G(\vec{x})$ and $T(\vec{x})$, we assume the following:

1. The target distribution is known.
2. The search points are independently chosen from the guess distribution.
3. The time required to move from guess point to guess point is negligible.

These assumptions describe the experimental case quite well after learning has taken place. Respectively, the subject is fast and successful, the correlations between eye positions (guesses) are weak in the sense that their length scale is $\lesssim 1°$ (Fig. 3), which is less than the length scale of the search, and the eye moves very quickly compared with time spent fixating. The assumption of uncorrelated steps likely fails as the length scale of the search decreases, that is, for $\sigma_T = 0.75$, where the problem stops being about search and more about moving the eye to a fixed, learned location.

Using these assumptions the general idea is to calculate the probability that a guess will be successful and optimize that to minimize the total number of guesses. The probability that a guess is successful for finding a target at location $\vec{x}$ is the integral of all points in the guess distribution that satisfy the range requirement:

$$P(\vec{x}) = \int_{\vec{C}(\vec{x})} d\vec{y} G(\vec{y}),$$

where $C(\vec{x})$ is the region where a target at position $\vec{x}$ would be found.

Then, the mean number of steps to find the target is one over the probability; for example, the mean number of times a die is rolled before finding a 1 is $6 = 1/p(1)$, where $p(1)$ is the probability of rolling a 1 or $1/6$. Averaging over the probability a target is actually at position $\vec{x}$, $T(\vec{x})$, leaves

$$\langle n \rangle = \int d\vec{x} \frac{T(\vec{x})}{\int_{\vec{C}(\vec{x})} d\vec{y} G(\vec{y})},$$

and the optimum occurs at

$$0 = \frac{\delta}{\delta G(\vec{y})} \left[\langle n \rangle + \alpha \left(1 - \int d\vec{t} G(\vec{t}) \right) \right],$$

where α is a Lagrange multiplier associated with normalization of G. Under reasonable existence and compactness assumptions, this optimization problem is solvable (1) and most generally satisfies

$$\frac{T(\vec{x})}{\left(\int_{\vec{C}(\vec{x})} d\vec{y} G(\vec{y}) \right)^2} \propto 1,$$

where, intuitively, the functional derivative of $1/G$ gives $1/G^2$. This has an interesting expansion when $C(\vec{x})$ is small enough that the integral over G can be approximated with the mean value theorem, leaving

$$G(\vec{x}) \propto \sqrt{T(\vec{x})}.$$

Optimum Guess Distribution for a 2D Gaussian Target Distribution

To interface with the main text, the target distribution is always a 2D Gaussian with a square window of $\pm R = 2°$ for finding the target. Here we will apply the search theory to calculate ideal values of the measured quantities.

Guess Distribution. For the case of a 2D Gaussian target distribution,

$$G(x,y) \propto \sqrt{\exp\left(-\frac{x^2+y^2}{2\sigma_T^2} \right)}$$
$$\propto \exp\left(-\frac{x^2+y^2}{2\left(\sqrt{2}\sigma_T\right)^2} \right).$$

In other words, for a Gaussian target distribution with SD σ_T, the optimum guess distribution is also a Gaussian but with SD $\sigma_G = \sqrt{2}\sigma_T$.

The square root rule holds only when the size of the search window is very small, but a low-order correction taking into account the window size has been derived elsewhere (1). The basic idea is that as the window size increases, the width of the guess distribution decreases because the searcher sees the tails for free when looking near the center. The general result for Gaussians and a search window of size $\pm R$ is that the optimum guess distribution is also a Gaussian with SD

$$\sigma_G = \sqrt{2}\sqrt{\sigma_T^2 - \frac{R^2}{\pi^2}}, \qquad \textbf{[S1]}$$

and that is the formula used to estimate the ideal behavior in the main text.

Mean Distance. The theory presented here predicts that the optimum guess distribution is a 2D Gaussian with the same center as the target distribution with a specific SD. In that case the mean distance from the center of the target distribution is

$$\langle d \rangle = \int dx \int dy \frac{\sqrt{x^2+y^2}}{2\pi\sigma_G^2} \exp\left(-\frac{x^2+y^2}{2\sigma_G^2} \right).$$

By switching to radial coordinates, the integral can be easily done to give

$$\langle d \rangle = \sqrt{\frac{\pi}{2}}\sigma_G. \qquad \textbf{[S2]}$$

Number of Steps. The case of interest is 2D, and that happens to be solvable in the sense that we can derive the mean number of steps to find the target in terms of the SD of the guess and target distributions. The average number of samples required to find the target is

$$\langle n \rangle = \frac{\sigma_G^2}{\sigma_T^2} \int dx \int dy \frac{e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma_T^2}}}{\int_{x-R}^{x+R} du \int_{y-R}^{y+R} dv e^{-\frac{1}{2}\frac{u^2+v^2}{\sigma_G^2}}}.$$

The double integral in the denominator represents the probability of contacting the target for a guess at position (u, v), and the numerator weights the probability that the target is actually at that location. Here the boundaries are taken to be at infinity because we will stay far away from them.

The integral in the denominator can be represented in terms of complementary error functions as

$$I_d = \frac{1}{4}\left[\operatorname{erf}\left(\frac{R+x}{\sqrt{2}\sigma_G}\right) - \operatorname{erf}\left(\frac{x-R}{\sqrt{2}\sigma_G}\right)\right] \times \left[\operatorname{erf}\left(\frac{R+y}{\sqrt{2}\sigma_G}\right) - \operatorname{erf}\left(\frac{y-R}{\sqrt{2}\sigma_G}\right)\right].$$

Each of the differences of erfs (one for x and one for y) can be approximated as a Gaussian by matching the zeroth and second moments. Those are

$$I_{d,x}^{[0]} = \operatorname{erf}\left(\frac{R}{\sqrt{2}\sigma_G}\right)$$

and

$$I_{d,x}^{[2]} = -\sqrt{\frac{2}{\pi}}\frac{R}{s^3} e^{-\frac{x^2}{2\sigma_G^2}}.$$

Then, by matching moments, I_d is well approximated as

$$I_d \approx \operatorname{erf}^2\left(\frac{R}{\sqrt{2}\sigma_G}\right) e^{-\frac{x^2+y^2}{2\sigma_H^2}},$$

where

$$\sigma_H^2 = \frac{\sigma_G^3}{R}\sqrt{\frac{\pi}{2}}\, erf\left(\frac{R}{\sqrt{2}\sigma_G}\right) e^{\frac{R^2}{2\sigma_G^2}}.$$

Finally, the double integral remaining for $\langle n \rangle$ is just a Gaussian integral:

$$\langle n \rangle \approx \frac{1}{2\pi\sigma_T^2} \operatorname{erf}^{-2}\left(\frac{R}{\sqrt{2}\sigma_G}\right) \times \iint dxdy \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_T^2} - \frac{1}{\sigma_H^2}\right)(x^2+y^2)\right].$$

The 2D Gaussian integral happens to be exactly solvable and is

$$\langle n \rangle \approx \operatorname{erf}^{-2}\left(\frac{R}{\sqrt{2}\sigma_G}\right) \frac{\sigma_H^2}{\sigma_H^2 - \sigma_T^2}. \qquad \textbf{[S3]}$$

This puts the average number of steps as a direct function of σ_G and the given parameters R and σ_T (Fig. S1).

The measurement of the mean number of steps is difficult from subject data because of the limited sample size. The agreement is reasonably good with subjects approaching the optimum number of steps (Table S1). Subjects do not actually reach the optimum, but they get within a few steps of it. This may either be because the behavior is good enough, or it may reflect errors due to the sequential effects that overemphasize looking near the last found target. In the actual case of the experiment with randomly placed targets, a tendency to go back to a recently rewarded location increases the number of samples.

Simulated Step Number. To verify the calculation for the number of steps and incorporate potential biases of the guess distribution, we ran direct simulations of the experiment. A step of simulation was as follows. For a given target distribution (Gaussians centered at zero without loss of generality), we chose a random target point. Then, from a given guess distribution (another Gaussian, centered either 0° or 1° away with fixed σ_G), we chose search points until the search was within $\pm 2°$ of the target in both the x and y directions. The number of samples was then tabulated, and its average and SE for each choice of the guess and target distributions estimated the average number of saccades required to find the target. Note that in Fig. S1 the data points agree nearly perfectly with the theory lines, although the approximations used to generate the theory lines start to break down for the smallest σ_G as expected because there was an expansion in σ_T/σ_G.

Reinforcement Learning Model

Reinforcement Learning. Participants' behavior in the Hidden Target Search Task is modeled within a reinforcement learning framework (2). In reinforcement learning (RL), an agent operates in a finite-state environment and takes actions that move it from one state to another, leading to states associated with reward. In the Hidden Target Search Task, the state space consists of the eye position in the 2D display. We discretize this eye position in units of 1°; allowing eye position to vary from −12° to +12°, the state space consists of 25×25 locations. At the start of the trial, the initial state of the model is at the fixation point in the center of the display, $(0, 0)$. The actions available to the model consist of saccades specifying relative movement of the eyes, also discretized. RL is concerned with discovering the action or action sequences that lead to a reward.

The simplest RL approach is to define a value function, $V(s)$, that specifies the expected reward associated with each state s. The value function might be implemented as a look-up table associating each of the 25×25 states with an expected reward. When the target is detected, the value associated with the current state is updated to reflect a rewarding event. The RL model should select an action, a, in state s that maximize the value $V(s+a)$. Because the eyes can in principle move from any position to any other position, decision making according to such a model is independent of the current eye position: At each decision point, the model should saccade to the most promising location on the screen.

This intuitive model is naive as an explanation of how people learn in two regards.

1. A look-up table encodes space as a set of distinct, nonoverlapping locations. In contrast, neural representations in visual cortex are coarse-coded: Neurons have broad, overlapping receptive fields. Consequently, any learning about one location will have a natural generalization gradient to nearby locations. We incorporate this notion of generalization via a kernel-based RL approach (3) in which the value function is represented by a look-up table, but instead of updating the value associated with the current state s, $V(s)$, all states q are updated with strength—or eligibility as it is called in the RL literature—proportional to

$$\exp(-\mu\|s-q\|), \qquad \textbf{[S4]}$$

where μ is a free parameter of the model characterizing the generalization gradient.
2. Individuals performing Hidden Target Search Task show a strong proximity bias that favors shorter saccades, as illustrated by the distribution of saccade vectors (shifts in position

from one fixation to the next) observed in our experiment (Fig. S2). This proximity bias could be an emergent consequence of some more fundamental cause, such as variability in saccade outcomes growing with saccade distance (4, 5) or representational inhomogeneities in the superior colliculus (6). The bias has been noted by other researchers and can override knowledge that subjects have about the expected location of a target (7). However, we model the bias directly by assuming a cost that grows with saccade distance, leading to an action selection rule in which the probability of making a relative saccade, a, from some state s is proportional to

$$P(a|s) \propto \exp\left(V(s+a) - \rho\|a\|\right), \qquad \textbf{[S5]}$$

where ρ is a free parameter that scales the cost. This bias transforms a task that has no inherent sequential structure (such as maze learning) into a sequential decision task—a task in which the model must plan sequences of actions to obtain a reward. Thus, the temporal difference (TD) learning paradigm (2) is appropriate for modeling learning and performance.With these two extensions, our model can be cast in terms of traditional TD value function learning, with an exploration policy specified by a slightly embellished version of Eq. **S5** in which we incorporate an exploration parameters β (for softmax action selection) and ε (for ε-greedy action selection):

$$P(a|s) \propto \exp\left(\beta V(s+a) - \rho\|a\|\right) + \epsilon. \qquad \textbf{[S6]}$$

Both forms of exploration were included because we were not certain a priori which form would better match participant behavior. Softmax exploration turned out to be more important, but ε-greedy exploration better accommodated occasional off-policy (random) actions. Our model included the standard TD trace-decay parameter λ, usually notated as TD(λ), discount parameter γ, and learning rate α, yielding the TD rule for updating the entire value function after each fixation:

$$\Delta V(q|s,s') = \alpha\left(r + \gamma V(s') - V(s)\right)\text{eligibility}(q), \qquad \textbf{[S7]}$$

where s is the current state (fixation) of the model, s' is the next state, r is the instantaneous reward (1 if target is found at s, 0 otherwise), q is an index over all states, and eligibility(q) is an additive eligibility trace associated with state q, updated following each saccade according to TD(λ) with the additional spatial blurring function of Eq. **S4**:

$$\Delta\text{eligibility}\,(q|s) = (\gamma\lambda - 1)\text{eligibility}\,(q) + \exp\left(-\mu\|s-q\|\right). \qquad \textbf{[S8]}$$

Bayesian interpretation of model. The model can be interpreted from a Bayesian perspective in which the value function specifies the log likelihood of reward given an action (saccade) and the proximity bias specifies log priors over actions given the current eye position. Via Eq. **S6**, actions are then selected from a posterior distribution conditioned on the model's current eye position and experience history.

Training the model. The model has in total seven free parameters: α, β, γ, ε, λ, μ, and ρ. For each session of each participant, we searched for the parameters that yielded the maximum likelihood fit to the fixations of the first 20 trials of the session. The likelihood is based on the joint probability of the fixation sequence over the first 20 trials, or, using the chain rule, the product of the probabilities of each fixation contingent on the history leading up to that fixation. Given a set of parameters, the model specifies this probability. Over participants, there was a mean of 473 fixations in the first 20 trials.

We used an L1 distance metric in both Eqs. **S6** and **S8**. We noted in the data that many participants tended to produce horizontal and vertical saccades, suggesting a city block metric to be more sensible as a measure of proximity of two screen locations. Confirming this intuition, we found that the fixation likelihoods were higher using an L1 than using an L2 metric.

To assess the validity of our assumption that the reward signal has spatial blurring, as embodied in μ, we constrained μ to a large value (meaning no blurring) and refit the model. Similarly, to assess the validity of our assumption of a saccade proximity bias, as embodied in the coefficient ρ, we constrained the coefficient to be zero and refit the model. In both cases, we found that removal of the assumption yielded a noticeable decrease in the goodness of fit to the fixation sequences, suggesting that these two assumptions are warranted.

Fig. S3 depicts a sequence of saccades from one participant on one trial, along with model predictions of where the participant will look. This example follows 10 earlier successes at locating the target. As the example shows, the model does a reasonable job of predicting the participant's specific eye movements.

Fig. S4 shows an example of the evolution of the value function across trials as estimated by the RL model for the specific fixation and reward sequence of one human participant.

Testing the model. Following training, in which model parameters are determined from a participant's fixation sequence, the parameterized model can be run in generative mode from a de novo state to simulate the participant performing the task over a session. At the start of the session, the value function is reset by assigning all entries the value zero. The model begins each trial with fixation at the origin, and each subsequent fixation is drawn from the distribution specified by Eq. **S6**. Given feedback—success or failure in locating the target—the eligibility trace and value function are updated (Eqs. **S7** and **S8**). This cycle repeats until the target is found or until the trial limit is reached. Because the limit during the experiments was defined by the passage of time (20 s) and the model operates in abstract simulation steps, we assumed the mean time per fixation for the model was the same as for its corresponding participant and terminated the trial when the simulation time limit was reached. Consequently, the model did not find the target on every trial. The specific target location chosen on each trial was the same as the location chosen for the corresponding participant on that trial.

To collect statistics on the model's performance, we used the maximum likelihood parameter estimate for each participant in each condition and ran 200 replications of the parameterized model in generative mode using the same target location sequence over trials as was shown to the participant. The replications differed from one another in stochastic action selection. Model statistics reported in the main text are means over the 200 replications, and either means or medians over the participants.

Individual differences in the emergent behavior of the model. Because only the first 20 trials were used for training, and because the criterion used to train the model (likelihood of a specific fixation sequence) is only loosely related to the various performance measures obtained from the model as it generated sequences of decisions (e.g., mean distance from target centroid), it is nontrivial that the model's generated behavior matches the participants'. Thus, we consider the simulation results presented in the main text to be emergent predictions of the model.

Nonetheless, the results are limited in that they merely show that the aggregate behavior of individuals corresponds to the behavior of the model, averaged over multiple instantiations. The results say little about how well the model can account for individual differences in performance. To explore how the model—when trained on fixation sequences of a particular individual—characterizes that individual's behavior, we examined two statistics of an individual's performance once behavior stabilized (trials 31–60): (i) mean distance of fixations to the target centroid and (ii) fixation spread.

These two measures are shown in aggregate across participants in Fig. 2 *A* and *B* of the main text. Each of these measures correlates well with the performance obtained by the model parameterized for the individual (distance: Spearman's $\rho = 0.40$, 0.71, and 0.53 for target spreads 0.75, 2.00, and 2.75, respectively; spread: Spearman's $\rho = 0.24$, 0.62, and 0.68). Thus, the model characterizes the relative search distributions of individuals.

In a further exploration of individual differences, we observed that some participants tended to begin a trial by making a sequence of fixations that gradually approached the target centroid, whereas others made a saccade to the centroid and then moved away on subsequent fixations. We characterized the initial trajectory by computing the mean distance to the target centroid of fixations 1–5 on trials 31–60 and computed the slope of the regression line. Negative and positive slopes indicate trajectories toward and away from the centroid, respectively. The Spearman correlation coefficient relating the model's predicted slope to the corresponding participant's slope was $\rho = .48$, indicating that the model does a fair job of describing aspects of strategic performance.

Sequential Dependencies. Based on observations of the RL model, we predicted that participants in the experiment would exhibit sequential dependencies—fixation behavior that was dependent on the recent history of fixations and reward. We showed in the main text that both model and participants showed reliable trial-to-trial dependencies. Fig. 3 shows that sequential dependencies extend back beyond the previous trial, and there is a decaying influence of recent experience. The value-function update (Eq. **S7**) is consistent with this decay, with exponential rate related to α. Thus, performance is never static, but continually adapts to the ongoing stream of experience. In both model and participants, sequential dependencies persist throughout the session; they are not merely a transient that occurs while the learner is still trying to ascertain the target distribution.

Sequential dependencies occur in the RL model because the value function is updated after every fixation, and with a constant learning rate the function does not converge (i.e., adjustments to the value function do not go to zero over time). Consequently, there is a cost in performance because the true target distribution is never exactly learned. More experience does not guarantee a better representation of the target distribution. In contrast, the ideal-observer theory is premised on perfect knowledge of the target distribution. It specifies the level of human performance that can be attained in a stationary environment.

Naturalistic environments are certainly nonstationary. In the presence of nonstationarity, there is a benefit of continued sensitivity to recent experience: When the environment changes, the agent will rapidly adapt. The trade-off between performing well in a fixed environment and rapidly learning a new environment naturally leads to sequential dependencies (8–10). It remains a final piece in the puzzle to develop an ideal-observer theory of search in environments with known nonstationary characteristics, combining both (1) and ideal-observer theories of change detection (11, 12).

Addressing this puzzle may also help tackle a longstanding question in visual search: What role does memory play? The role of memory in visual search has been debated (13–15), and the consensus is that the choice of the next saccade is informed by memory for a small number of previously attended locations. The value function in our RL model provides the substrate by which this memory could be implemented. According to the model, the memory is limited because (*i*) the value function encodes long-term memory as well as short-term experience, and the two are superimposed, (*ii*) the spatial generalization of reward (via the μ parameter) blurs the memory, and (*iii*) exploration (via the β parameter) weakens the guidance of memory on action selection.

1. Snider J (2011) Optimal random search for a single hidden target. *Phys Rev E Stat Nonlin Soft Matter Phys* 83(1 Pt 1):011105.
2. Sutton RS (1988) *Learning to Predict by the Methods of Temporal Differences* (Kluwer, Boston), pp 9–44.
3. Ormoneit D, Sen S (2002) Kernel-based reinforcement learning. *Mach Learn* 49:161–178.
4. Abrams RA, Meyer DE, Kornblum S (1989) Speed and accuracy of saccadic eye movements: characteristics of impulse variability in the oculomotor system. *J Exp Psychol Hum Percept Perform* 15(3):529–543.
5. Engbert R, Nuthmann A, Richter EM, Kliegl R (2005) SWIFT: A dynamical model of saccade generation during reading. *Psychol Rev* 112(4):777–813.
6. Sparks DL, Holland R, Guthrie BL (1976) Size and distribution of movement fields in the monkey superior colliculus. *Brain Res* 113(1):21–34.
7. Araujo C, Kowler E, Pavel M (2001) Eye movements during visual search: The costs of choosing the optimal path. *Vision Res* 41(25-26):3613–3625.
8. Mozer MC, Shettel M, Vecera S (2006) Top-down control of visual attention: A rational account. *Advances in Neural Information Processing Systems*, eds Weiss Y, Schoelkopf B, Platt J (MIT Press, Cambridge, MA), Vol. 18, pp 923–930.
9. Yu AJ, Cohen JD (2008) Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems*, eds Koller D, Schuurmans D, Bengio Y, Bottou L (MIT Press, Cambridge, MA), Vol 21, pp 1873–1880.
10. Wilder MH, Jones M, Mozer MC (2010) Sequential effects reflect parallel learning of multiple environmental regularities. *Advances in Neural Information Processing Systems*, eds Bengio Y, Schuurmans D, Lafferty J, Williams C, Culotta A (Neural Information Processing Systems Foundation, La Jolla, CA), Vol. 22.
11. Adams RP, MacKay DJ (2007) Bayesian online changepoint detection. arXiv: 0710.3742v1 [stat.ML].
12. Brown SD, Steyvers M (2009) Detecting and predicting changes. *Cognit Psychol* 58(1): 49–67.
13. Beck MR, Peterson MS, Vomela M (2006) Memory for where, but not what, is used during visual search. *J Exp Psychol Hum Percept Perform* 32(2):235–250.
14. Horowitz TS, Wolfe JM (1998) Visual search has no memory. *Nature* 394(6693): 575–577.
15. Horowitz TS, Wolfe JM (2001) Search for multiple targets: Remember the targets, forget the search. *Percept Psychophys* 63(2):272–285.

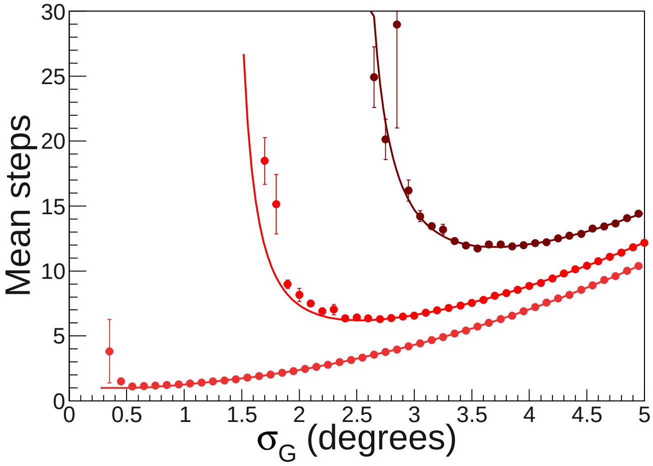

Fig. S1. The theoretical mean number of steps required to find the target for $\sigma_T = 0.75, 2.00, \text{and} 2.75$, bottom line to top, respectively. The lines are theoretical values, and the points are calculated with a direct simulation of the problem (1 million trials per point). The minimum point agrees well with the calculated values.

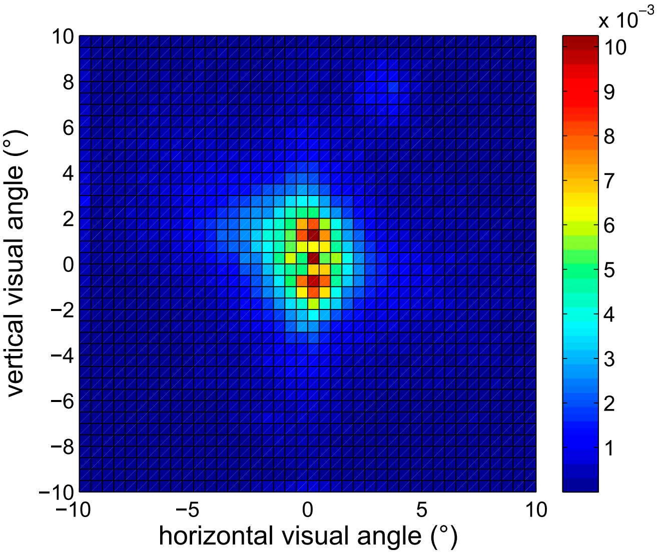

Fig. S2. Histogram of the saccade vector (in degrees visual angle), computed across participants, sessions, trials, and fixations within a trial.

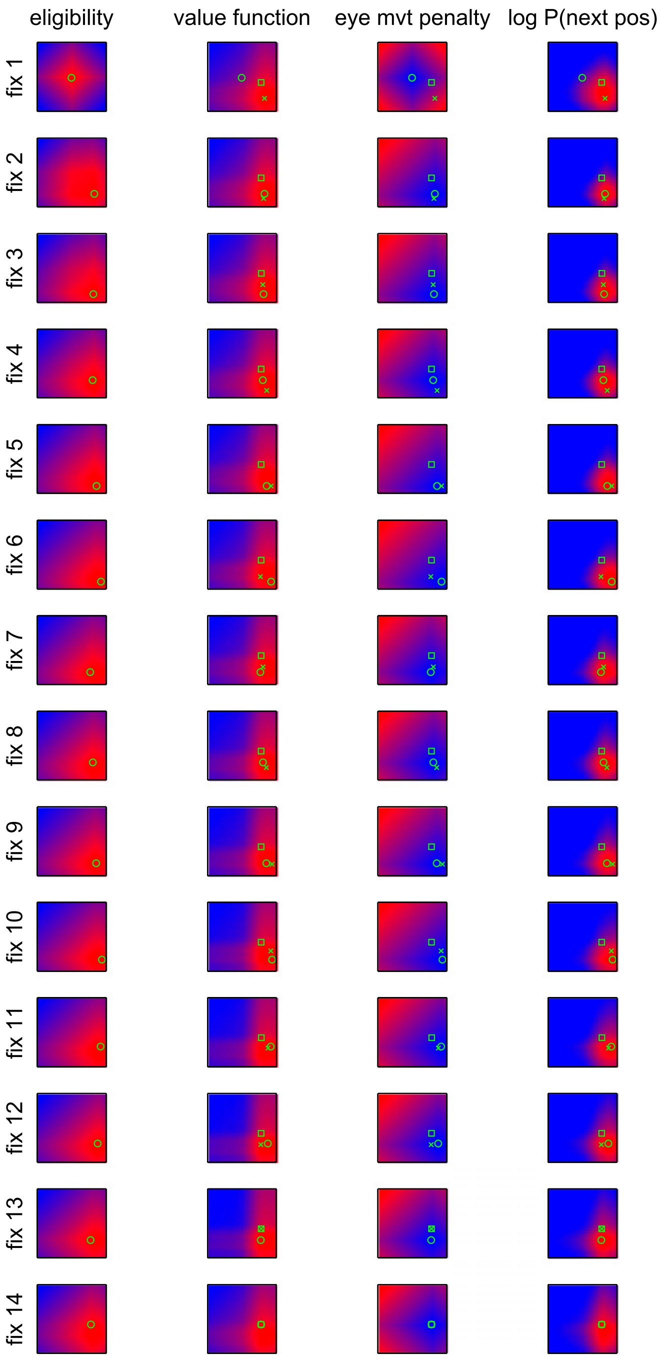

Fig. S3. Simulation showing model behavior on one trial for one participant. Each row depicts a single fixation within the trial. The four columns depict the eligibility trace, the value function, the penalty associated with an eye movement based on distance from the current location, and the model's probability of selecting a saccade destination. The open square in each panel indicates the target location for that trial, which is fixed. The open circle indicates the current fixation (the trial starts with fixation at the center of the screen). The X indicates the participant's next saccade destination. In all panels, values are indicated by coloring, where low values are blue and high values are red. For example, in the final column, note that the actual fixation lies in the red region—the region where the model predicts the next saccade to occur. The model correctly predicts that the participant will move their eyes to neighborhood of the target centroid in the first saccade, and then will make short fixations in that neighborhood.

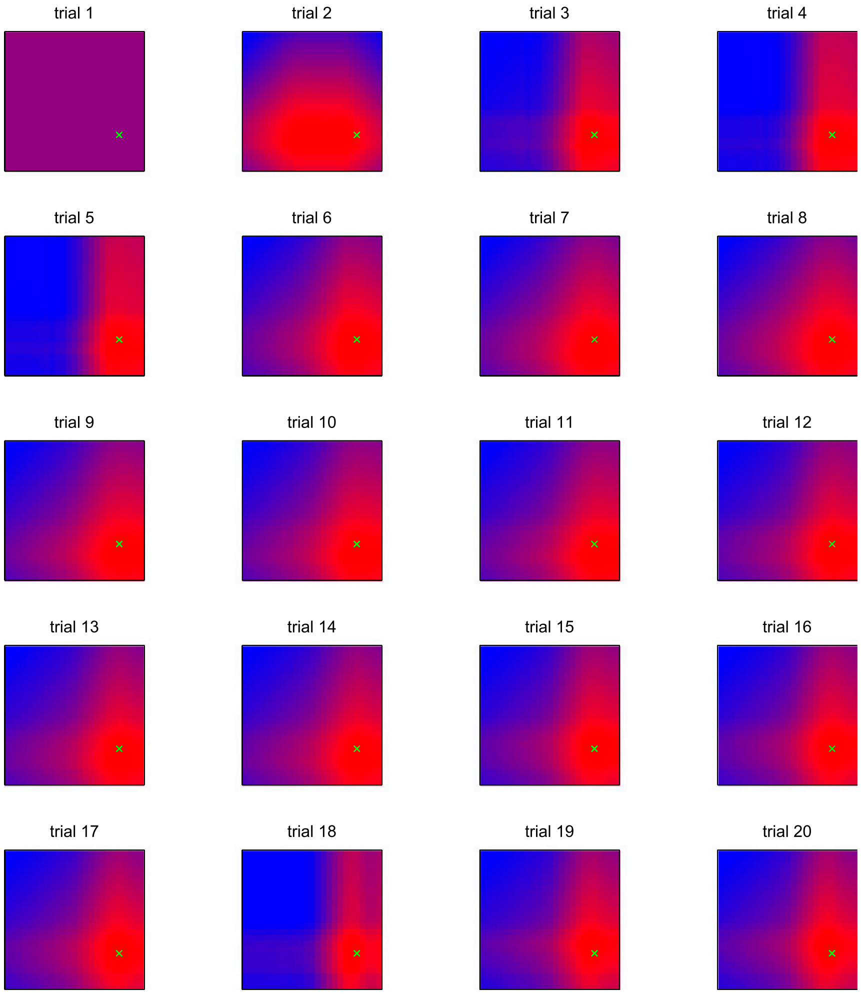

Fig. S4. Value function learning: The RL model's estimate of the value function given the fixation sequence of a particular individual in the experiment. Each panel depicts the value function at the start of a trial for trials 1–20. The color scale ranges from blue to red, for low to high reward expectation. The green X indicates the target centroid. The target neighborhood is rapidly identified, although the value distribution shifts slightly from trial to trial based on recent experience.

Table S1. Mean steps to find the target

σ_T,°	$\langle n \rangle_{\text{found}}$	$\langle n \rangle_{\text{all}}$	$\text{median}_{\text{found}}$	$\text{median}_{\text{all}}$	$\langle n \rangle_{\text{theory}}$
0.75	2.6 (3)	2.6 (3)	2.0 (3)	2.0 (3)	1.01
2.00	8 (2)	18 (4)	6 (2)	6 (4)	6.26
2.75	17 (3)	22 (4)	13 (3)	15 (4)	11.86

Numbers in parentheses are the standard error in the last digit.

Table S2. Bias toward the center and its effect on the measured quantities

σ_T,°	Measured bias, °	$\sigma_{G,\text{theory}}$, °	$\langle n \rangle_{\text{theory}}$
0.75	0.6 (3)	0.85 (5)	1.58 (1)
2.00	1.2 (3)	2.7 (5)	7.01 (2)
2.75	1.0 (5)	3.85 (5)	12.63 (3)

Positive bias is toward the center. Numbers in parentheses are the standard error in the last digit.