Learning where to look: An empirical, computational and theoretical account of hidden target search performance

Leanne Chukoskie1, Joseph Snider1, Michael C Mozer1,2, Richard J Krauzlis3 Terrence J Sejnowski1

1. Institute for Neural Computation, University of California at San Diego, La Jolla, CA 920932. Department of Computer Science and Institute of Cognitive Science, University of Colorado, Boulder, CO 80309 3. Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD 20892-4454. Howard Hughes Medical Institute, Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037 5. Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92039

Submitted to Proceedings of the National Academy of Sciences of the United States of America

We choose 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. Yet, remarkably little is known about how individuals learn to direct their gaze given the current context and task. To address this gap in our understanding of search behavior, we designed a new task in which participants search a novel scene for a target whose location was drawn stochastically on each trial from a fixed prior distribution. To focus on how participants learned this distribution, we made the target invisible on a blank screen, and the participants were rewarded when they fixated on the 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.

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 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). Ironically, 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 colleagues (23) have replicated this result for the different tasks of visual search and image memorization. However Wolfe and colleagues (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. 1A).

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 eye movement. A recent review and commentary from

Reserved for Publication Footnotes

www.pnas.org — — PNAS | Issue Date | Volume | Issue Number | 1—77
Wolfe and colleagues explore the notion of "semantic guidance" in complex, naturalistic scenes (28) 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. 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 (29). Other work has cast foraging behavior in terms of optimal search (30). What distinguishes foraging from visual search is that visual search tasks have visible cues that drive search, in addition to contextual information that specifies probable target location. In order 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 reinforcement-learning framework has become widely accepted for modeling performance in tasks involving a series of movements leading to reward (31, 32). In addition, for organisms across many levels of complexity, reinforcement learning has been shown to be an excellent framework to consider adaptive behavior in complex and changing environments (33, 34). Here, we describe performance in our task in terms of a reinforcement-
learning (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 (26) and its surroundings were available to guide the search. 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 new search task in which participants were rewarded for finding a hidden target, similar to the scenario encountered by a foraging animal (Fig. 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 standard deviation, respectively) that was held constant throughout a session (see Fig. 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. 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. 1C. 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. 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. 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 1).

A measure of search spread is the standard deviation of the set of fixations in a trial. The search spread was initially broad and narrowed as the session progressed, as shown in Fig. 2B 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. 2C).

Humans Approach Ideal Observer Performance. We now consider the behavior of participants once performance stabilized. Taking trials 31–60 to reflect asymptotic behavior, we examined the efficiency of human search in comparison to a theoretical optimum. An ideal observer was derived for the HTST 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 Figs. 2A–C mark ideal observer performance. Ideal search performance requires a distribution of planned fixation "guesses" that is 2° broader than the target distribution itself (35,36). As seen in Figs. 2B,C, the performance of participants hovered around this ideal search distribution after about a dozen trials. In Fig. 2A, the mean for the human data from trials 31–60 trend higher than the theory suggests, but the theory presumes stationarity of the target distribution. However, individuals must be responsive to nonstationarities in natural environments and this responsivity yields an increase in uncertainty (37) consistent with observed human performance. Across different target distribution spreads, the ideal-observer statistics qualitatively matched those of the human participants (Table 1), and the quantitative match was excellent for 2.00° and 2.75°.

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. Reinforcement-learning 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. S4), 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 (31), which are particularly appropriate for Markovian tasks in which sequences of actions lead to reward (see SI Materials and Methods for 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 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. 2A, the RL model approaches, but does not reach, the theoretical asymptote. Like the human participants, 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 (see SI Materials and Methods). Because the model had emergent dynamics and the data used to fit the model were quite different in nature from the statistics derived from the model, the observed consistency between participants and simulation suggests that the model is more than an existence proof of a mechanism, but could provide insight into the biological mechanisms of learning (41).

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. 3A), as predicted by the model (Fig. 3B) and explained some of the trial-to-trial variability in performance (Fig. 2 and SI Materials and Methods). Participants were biased to start the next trial's search near found target locations from recent trials. The influence of previous trials decreases exponentially, with the previous two or possibly three trials back possibly influencing the current trial's saccade choice (Fig. 3C). This exponential damping of previous trials' influence is approximated by the memoryless case (35), 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. 4). While individual trials look like Levy 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 Levy like, but separates into two distinct length scales like the intermittent search popularized by Bénichou (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. 4 is similar to that found in other demanding visual search tasks (35), as well as intermittent foraging by a wide range of animals (30, 43).

Discussion

Human search performance can be put into the more general context of animal foraging, which has close connections with reinforcement-learning models (34) 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). Our experimental results on how the distribution of hidden tar-
from different foraging-relevant modalities to encode the value of possible target locations and keep a running estimate of the most likely search targets to show that area LIP neurons differentiate between targets and distractors. NMDA and the promise of greater compensation, but no visual feedback. If after 20 seconds of search the target was not acquired, the trial ended with no tone.

The resulting search performance was well described by a reinforcement-learning model similar to that used previously for examining 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.

**Methods**

We defined a spatial region of an image as salient by associating it with underlying the learning and performance of the hidden target task, which may be shared with other search behaviors. 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 use 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 reinforcement-learning (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 saccade targets are generated.

Naturalistic environments are non-stationary. 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. 3) are a signature of this sensitivity (37, 46, 47). 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 the participants as a group, and also captured individual idiosyncrasies (see SI Materials and Methods). Sensitivity to non-stationary 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.

**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 appear to carry signatures of task components required for successful visual search. The lateral intraparietal area (LIP) and the superior colliculus 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 (48, 49).

Recently, Bisley and colleagues 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 (50). Area LIP neurons integrate information from different foraging-relevant modalities to encode the value associated with a movement to a particular target (51, 52) The neural mechanisms serving patch stay-leave foraging decisions have recently been characterized in a simplified visual choice task (53), providing a scheme for investigations of precisely how prior information and other task demands mix with visual information available in the scene. Sub-threshold microstimulation in area LIP (54) or the SC (55) also biases the selection of saccades toward the target in a complex visual environment. These results further 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 (56). Hikosaka and colleagues 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 (SNr) (5757). As the cells carrying this stable value information appear to project preferentially to the SC, these signals are well-placed to influence saccade choices through a fast and evolutionarily conserved reward system for controlling orienting behavior, but no visual feedback. If after 20 seconds of search the target was not acquired, the trial ended with no tone. More details about observers, stimuli, equipment, and procedure are given in SI Materials and Methods.

ACKNOWLEDGEMENTS: This work was supported in part by NSF grant #BES-0542013 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center (LC, JS, MM, TS), a Blasker Rose-Miah grant from the San Diego Foundation (LC), ONR MURI Award No.: N00014-10-1-0072 (JS, TS), and Howard Hughes Medical Institute (TS.).


