Cognitive Science 33 (2009) 530–541

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01021.x

Exploration and Exploitation During Sequential Search

Gregory Dam,[a,b] Konrad Körding[a,b]

[a]*Department of Physiology, Feinberg School of Medicine, Northwestern University*
[b]*Rehabilitation Institute of Chicago*

Received 15 April 2008; received in revised form 20 October 2008; accepted 13 November 2008

Abstract

When we learn how to throw darts we adjust how we throw based on where the darts stick. Much of skill learning is computationally similar in that we learn using feedback obtained after the completion of individual actions. We can formalize such tasks as a search problem; among the set of all possible actions, find the action that leads to the highest reward. In such cases our actions have two objectives: we want to best utilize what we already know (exploitation), but we also want to learn to be more successful in the future (exploration). Here we tested how participants learn movement trajectories where feedback is provided as a monetary reward that depends on the chosen trajectory. We mathematically derived the optimal search policy for our experiment using decision theory. The search behavior of participants is well predicted by an ideal searcher model that optimally combines exploration and exploitation.

Keywords: Human search behavior; Neuroeconomics; Skill acquisition; Decision making; Motor control; Mathematical modeling

1. Introduction

When we forage for food, search for a good price in a market, or learn a new motor skill, such as dart throwing, we learn from the outcomes of our actions. In such cases, we typically have uncertainty about the values and probabilities of the potential outcomes of our decisions. The challenge in such tasks can be conceived as a search problem; among the set of all possible actions, find those that result in desirable outcomes. The goal is to make a sequence of decisions, each one informed by previous decisions, with the objective of maximizing expected rewards.

During search we can receive immediate rewards for successful actions, and we can also improve our knowledge about the relationship between possible actions and their outcomes.

Correspondence should be sent to Gregory Dam, Rehabilitation Institute of Chicago, Chicago, IL 60611. E-mail: g-dam@northwestern.edu

This implies that decisions should be driven by two objectives, which are learning about the outcomes and receiving rewards. Actions that lead to large rewards are desirable. However, we may be willing to sacrifice an immediate reward for greater future rewards. For this reason, actions that reveal information are also desirable.

Much progress has been made in the field of reinforcement learning toward formalizing the two computational goals of exploration and exploitation (Kaelbling, Littman, & Moore, 1996; Sutton & Barto, 1998). Exploration is the process of choosing actions with the objective of learning about the environment we are interacting with. Exploitation, on the other hand, is the process of using previously obtained information to acquire rewards. Optimal strategies will combine these two objectives appropriately and result in behavior that is both informative and rewarding. For example, when there is much uncertainty about the environment, searching should be dominated by the goal of exploring. On the other hand, when uncertainty is low, there should be a bias toward more exploitation (Daw, Niv, & Dayan, 2005).

Search has been a topic of much interest for a variety of disciplines, including computer science (Kaelbling et al., 1996; Sutton & Barto, 1998), economics (Kohn & Shavell, 1974; Weitzman, 1979) and behavioral ecology (Bartumeus & Levin, 2008; Krebs, Kacelnik, & Taylor, 1978; Montague, Dayan, Person, & Sejnowski, 1995). In the domain of computer science, impressive progress has been made toward developing efficient search algorithms. The way animals and people search, however, is still poorly understood. This may be because the behavior of biological organisms is driven by a multitude of goals and objectives that may vary over time. Various studies address how factors such as changing costs, task demands, or the environment affect search behaviors (Mayntz, Raubenheimer, Salomon, Toft, & Simpson, 2005; Pyke, 1984). While search behavior has been predicted successfully in some specialized cases (Najemnik & Geisler, 2005; Viswanathan, Buldyrev, Havlin, Luz, Raposo, & Stanley, 1999), understanding how people balance the goals of exploration and exploitation is still an important problem (Cohen, McClure, & Yu, 2007).

To understand how people solve search problems we use techniques from the study of sensorimotor integration. Experimental measures of human sensorimotor performance have often been predicted well by models of optimal behavior (Flash & Hogan, 1985; Harris & Wolpert, 1998; Todorov & Jordan, 2002). Sensorimotor tasks often implicitly define a reward function for the participant that is intuitively clear, such as accurately landing on a displayed target. This leads to small variations across trials for each participant and behavior that is roughly conserved across all participants. The case of motor learning can often be reduced to a search problem. It can be understood as a search through a high dimensional motor space for a movement policy that yields desired results.

Here we use a movement experiment to test which strategies people use when they search. In our experiment participants perform a one-dimensional movement search task, where search decisions are drawn from a continuous set of possible actions. After every search, feedback is provided in the form of a monetary reward. We measure how search decisions are affected by past rewards. We chose the search task so that we are able to mathematically derive the search policy that combines exploration and exploitation in an optimal way. Human behavior is remarkably close to the behavior predicted by this ideal search policy.

2. Method

2.1. Participants

Eight right-handed participants (four females), with a mean age of 27.4 were paid a participation stipend proportional to their search performance. The average participation stipend was 24.3 dollars.

2.2. Materials

Participants were seated in a chair facing a computer screen. They were asked to draw trajectories with their dominant hand using the stylus of a PHANToM Premium 1.0 haptic robot (SensAble Technologies, Inc., Woburn, MA). Trajectories were made by sliding the tip of the robot's stylus along the surface of a desk (Fig. 1A). With the stylus, participants controlled the position of a cursor on the computer screen.

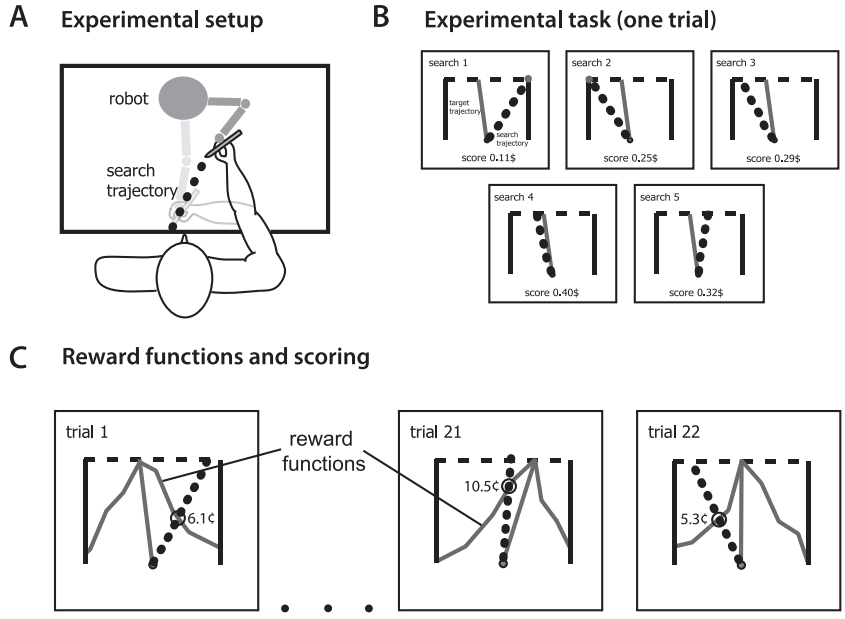

Fig. 1. Experimental paradigm and reward functions. (A) The experimental setup. (B) A typical trial consisting of five searches. The dotted lines represent participants' search trajectories. The gray lines are the target trajectories, which are not shown during the experiment. (C) Before each trial, a new target trajectory and reward function are defined and remain unchanged until the next trial (i.e., after five searches). The target and reward function are not shown during the experiment. Monetary rewards are computed by evaluating the intersection of each search trajectory with the reward function displayed as open circles. The figure also illustrates how the form of the reward function is varied from trial to trial.

2.3. Procedure

Participants completed 200 search trials. During each trial, five sequential hand movements in the form of straight planar trajectories were measured with a robotic manipulandum (Fig. 1A), for a total of 1,000 movements per experiment. Each drawn trajectory represented a single search attempt. Participants were instructed to decide on the direction of their trajectories so as to best approximate a hidden target trajectory (Fig. 1B). All search trajectories started at the same position and ended when the cursor reached a displayed line. The relevant decision made for each search attempt was choosing the direction of the trajectory within the confines of a bounded movement space. Taken trajectories were displayed as a trace of dots marking the path of the cursor. During the first two searches, participants were forced to sample near the bounds of the search interval. First, searching at the right- and then to the leftmost extremes of the search space. Participants were free to choose the direction of the subsequent three search trajectories. Forcing the first two searches reveals critical information about the values at the boundaries of the search space that might otherwise be ignored if it were simply displayed.

At the completion of each search trajectory the participant was shown a monetary reward. Participants received the highest monetary reward of the five searches for each trial, which were summed to determine the participation stipend (Fig. 1C). The amount of the monetary reward was a function of how closely the direction of the drawn trajectory matched the direction of the target trajectory. All target trajectories were straight planar paths that varied only in their direction, which was chosen for each trial from a uniform distribution (Fig. 1B). During each search trajectory the horizontal and vertical positions of the cursor were recorded at an approximate sampling frequency of 250 Hz. Since participants' trajectories are less than perfectly straight, a line was fitted to the sampled positions using a least-squares method to determine the intended direction of each movement. Monetary rewards were computed using a unimodal reward function with a maximum value defined at the exact direction of the target trajectory. Both the target and the reward function were held constant for each trial (i.e., five searches). Trajectories that deviated from the target direction elicited a less than maximum monetary reward defined by the reward function, which is strictly decreasing on both sides of the maximum. The magnitude of the reward was determined by interpolating the intersection between the line fitted to the subject's movement and the reward function (see Fig. 1C).

In order to obtain independent samples of search decisions from participants, we varied two aspects of the reward function at the beginning of each trial. First, the maximum achievable score was drawn randomly from monetary amounts ranging from $0.05 to $0.45 in five cent increments. Second, the shape of the reward function was varied in a parameterized fashion for each trial, maintaining the constraint that they all be unimodal (Fig. 1C). The purpose of varying the form of the reward function is to prevent participants from developing search strategies that would result from learning a static reward function. For example, participants may assume a strategy for choosing search locations that are proportional to an error calculated as the deviation of the current score from a known maximum possible score.

3. The ideal searcher

For some single dimensional search problems it is possible to derive the optimal search policy (Kiefer, 1953). More specifically, our derivation assumes that the reward function is strictly concave with a stationary target and that no additional information about the reward function can be used (e.g., no use of gradient information). A consequence of unimodal reward functions is that rewards must be strictly decreasing with increasing distance from the maximum. This implies that if we have two searches at a and b, where the ordering of the locations is $a < b$ and the ordering of the values is $f(a) > f(b)$, then the maximum cannot be at any location $x > b$. In subsequent searches the interval that must contain the maximum will be limited and parts of the search space may be eliminated (''dropped'') from future searching. We explicitly assume that the searcher cannot use any other information apart from knowing the locations of past searches and the interval that contains the maximum. Learning in this case is the process of dropping intervals in order to reduce the size of the interval containing the maximum and consequently reduce uncertainty about the target.

For every new search the relevant gained piece of information is whether the search results in a reward higher than every previous search. The information obtained from each search can be represented with a single bit. We assume, without loss of generality, that the first search (forced to be at the left end of the search interval, 0) yielded a smaller reward than the second search (forced to be at the right end, 1; see Method). If this is not the case we can simply switch the coordinate systems to obtain an analogous problem. The third search will only yield one additional bit of information to the searcher, which is whether the search yielded a higher or lower reward than the second. The fourth search, similarly, will only reveal one more bit of information, which is whether the fourth search reward was higher than the maximum reward of the first three searches.

The optimal strategy can thus be formulated by a policy that maps the set of bits that have been acquired to the next search position. A single parameter α is used to define the best third search location. Two parameters β_1 and β_2 define the fourth search in the case of either $f(\alpha) < f(1)$ or the case $f(\alpha) \geq f(1)$, respectively. Four parameters γ_{11}, γ_{12}, γ_{21}, γ_{22} define the locations after searching at β_1 or β_2, and determining whether the fourth search procures the largest reward (γ_{x1}, γ_{x2}, respectively). These seven parameters ($\theta := \{\alpha, \beta_1, \beta_2, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}\}$), fully describe the search decisions of an optimal agent. We can represent the search policy using a decision tree (Fig. 2). Each search position is associated with a node of the tree. Each edge in the graph has an associated probability of following the edge, which is the probability of dropping or not dropping an interval, respectively. Four possible paths through the tree exist, each representing a full set of search prescriptions for a search trial in our experiment. Each path is chosen conditional on the search reward history.

The best search policy is defined by seven parameters (see Fig. 2) each representing where to search given a particular reward history. The optimization of these parameters is a function of both their associated reward losses and the transition probabilities through the graph, which are in turn a function of the parameter values themselves. The optimal policy will lead to the smallest loss of reward and thus the highest overall payoff. Any complete search sequence can be characterized as a walk through the decision tree (Fig. 2). The

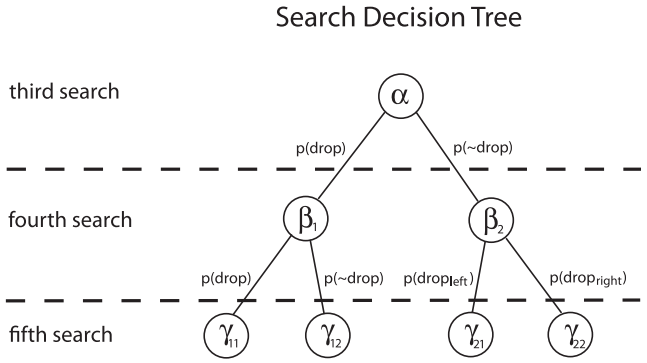

Fig. 2. Search decision tree. Each node represents a search decision. Each edge has associated with it the probability of transition from its parent node. The graph illustrates that there are four possible paths through the tree that depend on search results. Note that after the third search (alpha) the only possible outcomes are to drop or not drop an interval. When an interval is not dropped after the third search, then an interval must be dropped after the fourth search. In this case there are two possibilities: (a) dropping the far left interval or (b) dropping the far right interval.

expected loss can be calculated using decision theory by weighing the loss of ending at each of the leaves (situation after the last search) with the probability of ending up at that leaf.

The best policy is then the one that minimizes the total expected loss of the search policy:

$$\hat{\theta} = \arg\min_{\theta} \sum_{all\ paths} [Loss(path|\theta) \cdot P(path)|\theta)] \tag{1}$$

The probability of traversing a given path can be calculated by multiplying the probabilities of all the component edges of the path, where the transition probability associated with each edge is uniquely defined as a function of previous search locations. For example, the probability of dropping an interval after the third search is calculated as follows:

$$P(drop\ after\ \alpha\ search) = 1 - \frac{1+\alpha}{2} \tag{2}$$

To calculate the average loss associated with each search decision we assume unimodality and, therefore, payoffs decrease with increasing distance to the target and obtain:

$$Loss\big(path|\theta_{path}\big) = \arg\min_{\theta} \sum_{\theta_k \in path}^{K} \int_0^1 \big|\theta_k - x_{target}\big| \cdot P(x_{target}) dx_{target} \tag{3}$$

To find the optimal search policy we need to calculate the search policy with the lowest expected loss (Equation 1). To do so we implemented Equation 1 in MATLAB and then used the fminsearch function with multiple restarts to find a minimum. The optimal search

locations are computed to be as follows: $\alpha = .61$, $\beta_1 = .42$, $\beta_2 = .83$. The fifth search locations (γ_{11}, γ_{12}, γ_{21}, γ_{22}) are in the middle of the largest remaining interval.

3.1. Exploration and exploitation

Kiefer (1953) analytically derived the optimal exploration algorithm for searching for the extrema of unimodal functions. The goal of this algorithm is to generate intervals that are subsets of the entire search interval and that are guaranteed to contain the maximum. It is thus only aimed at reducing uncertainty (exploration) and not aimed at obtaining rewards. Applying this algorithm to our task, the optimal exploration searches are (Fig. 3):

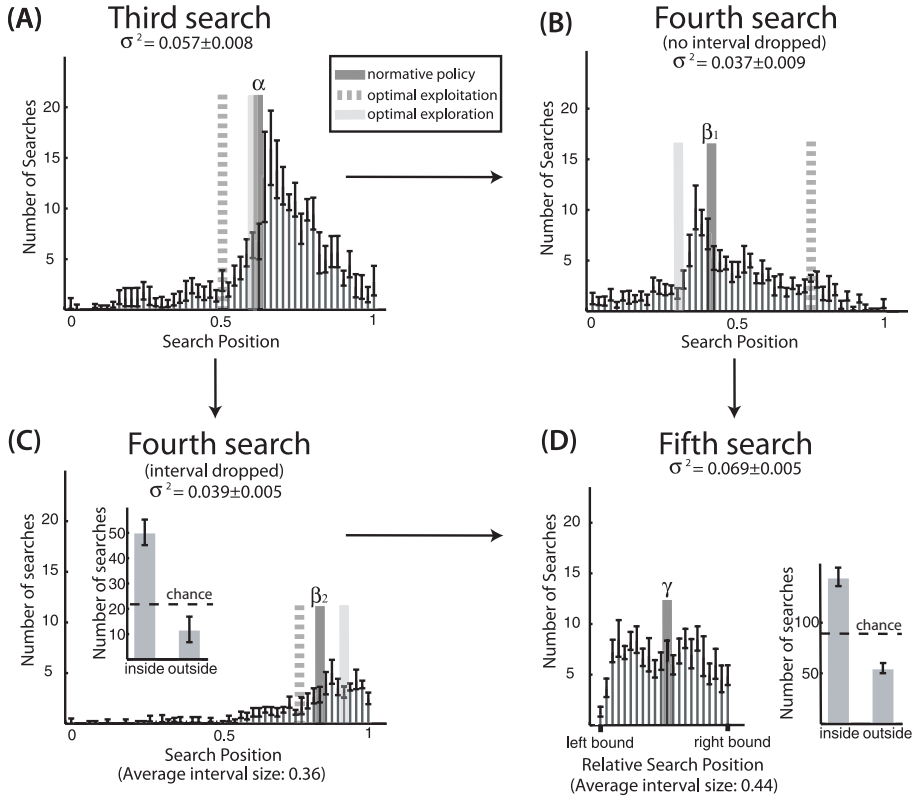

Fig. 3. Search behavior. Histograms of the average search location across all participants with standard error bars. Participants show a strong tendency to search at particular locations. These locations match closely to the search locations prescribed by the ideal policy. Also illustrated are the optimal exploitation and exploration strategies for the third and fourth searches. The gray bar graph insets in (C) and (D) show how often participants searched within the correct nondropped interval. All histograms represent the generalized case where the second search reward is greater than the first reward. In reality, this is true of roughly half of the search trials. In order to better display search behavior, all trials where the first search reward is larger than the second are graphed as (1-search location). In the unprocessed data, the search locations are distinctly bimodal for the third and fourth searches.

3rd = .60×(largest remaining interval), 4th = .66×(largest remaining interval), and 5th = .5×(largest remaining interval). A pure explorative strategy aims at having a small expected remaining interval. For the ideal searcher (as well as the subjects) the highest reward obtained is what matters. In Kiefer's case only the size of the remaining interval matters (for predictions see Fig. 3).

A pure exploitation strategy, on the other hand, has the goal of maximizing the expected reward on every single search. Under these reward conditions, a pure exploitive strategy would search at locations that minimize their distance to all possible target locations, which is in the middle of the largest nondropped interval (Fig. 3).

4. Results

Here we compare the behavior of human participants with the behavior of the ideal searcher. In many cases it has been found that human behavior is close to optimal and thus humans might behave as predicted by the ideal searcher. However, human participants may be limited in their capacity to search and could only be able to behave as predicted by one of the suboptimal strategies of exploration or exploitation. Our experiment is aimed at characterizing human search behavior within this framework.

We studied the way people search by analyzing the histograms of search positions as a function of previous search outcomes (Fig. 3). Note the strong peak in the histogram of the third search, indicating that human search behavior is relatively well conserved across participants. The modes of the histograms appear to be close to the search locations prescribed by the ideal search policy. To quantify this observation we compared the means of search positions across all participants and all searches to the theoretically derived search strategies considered above. The mean third search position for the eight participants was .67 (SD = .25) and is significantly different from the ideal searcher position, $t(7) = 9.29$, $p < .001$, the optimal exploration strategy, $t(7) = 3.92$, $p < .001$, and the optimal exploitation strategy, $t(7) = 3.39$, $p = .011$. A significance test failed to reject the null hypothesis of no difference between the means of the fourth search when an interval may be dropped (M = .81, SD = .26) and the ideal searcher $t(7) = -1.54$, $p = .16$. However, in this case the means are significantly different from both the exploration strategy, $t(7) = -5.16$, $p < .001$ and the exploitation strategy, $t(7) = 4.80$, $p < .01$. The fourth search in the no-drop case has a mean of .45 (SD = .27), not statistically different from the ideal searcher, $t(7) = 1.86$, $p = .10$, but distinct from the exploration strategy, $t(7) = 7.39$, $p < .001$, and the exploitation strategy $t(7) = -18.42$, $p < .001$. The histograms in Fig. 3 suggest that the human behavior tested in our experiment is largely consistent with the ideal searcher strategy. The mean search locations of participants were significantly different from both the exploration and exploitation strategies and failed to show a difference from the ideal search policy in all but one case.

We generally find that participants tend not to search within dropped intervals. The gray bar graphs in Fig. 3 show how often participants typically search in the correct intervals. Averaged across all participants, 81 ± 3% of the fourth searches and 72 ± 2% of the fifth

searches were in the correct interval. Searches within the correct interval are well above chance, which are 36% and 44% for the fourth and fifth searches, respectively. While participants clearly utilize information about the possibility of dropping intervals, their decisions may be affected by factors such as limited memory and motor noise.

As participants are only allowed five searches, information gained during the last search is not helpful and the ideal searcher is thus identical to the pure exploitative strategy. Both models predict a search in the middle of the largest remaining interval. However, we rather observe searches that are uniformly distributed within the largest remaining interval (Fig. 3). During the last search, the remaining interval is much smaller and thus motor noise or a lack of precision in the memory of the positions of past searches will have a more pronounced influence.

We characterize the efficiency of participants' search strategies by comparing the rewards they obtained with the predictions of various theories (Fig. 4). Simulations for the pure exploitation and pure exploration strategies result in lower average maximum reward on the last trial. A simulation of the modes of participants' search locations more closely matches the maximum rewards obtained from the ideal search policy simulation. The rewards participants actually obtained during the experiment are slightly lower than the simulated ideal rewards while being close to the achievable optimum. This finding relates to published findings that describe that human movement strategies often lead to rewards that are close to optimal (Maloney, Trommershäuser, & Landy, 2006; Trommershauser, Maloney, & Landy,

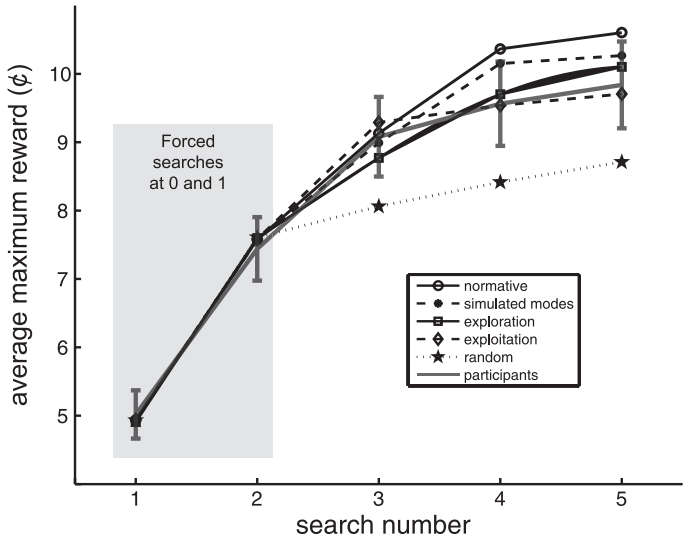

Fig. 4. Simulated rewards. Simulated rewards of the ideal strategy, pure exploitation strategy, pure exploration strategy, modes of the participants' search locations, and random search are graphed. Participants' average reward data with standard error bars are shown for comparison. Note that the simulated participants' modes yield the closest match to the rewards of the simulated ideal policy. Final search rewards for the ideal strategy are on average 0.50 and 0.89 cents higher than the exploration and exploitation strategies, respectively.

2003). In our task, participants show very efficient movement decision making even when there are considerable cognitive demands.

Even though people have clear tendencies to search at particular locations, there is considerable variance in search decisions. Participants are asked to make search decisions by performing movement trajectories. It has been generally found that people exhibit some motor noise when making movement trajectories, even straight reaching movements (Harris & Wolpert, 1998). Although this movement imprecision will account for some of the variance in search behavior, it would at best explain only a small percentage of the variance observed. Other limitations in people's searching in our task must be considered. Each participant performed one thousand arm movements over approximately 2 h of experimentation, which presents high demands on memory and attention. Limitations in these could account for some of the additional variance not explained by motor noise. For example, not all the relevant taken trajectory directions and their corresponding rewards will be accurately recalled whenever a new decision is made. Furthermore, there may be some aspect of the decision process itself that results in the observed deviation from optimal. Perhaps participants may at times engage in random exploration, which is shown to be optimal under certain conditions (Kaelbling et al., 1996). Or perhaps participants' utility functions for money have a nonlinear effect on their decisions as described in prospect theory (Kahneman & Tversky, 1979).

5. Discussion

To understand how people search along a single dimension we have analyzed their strategies using a motor learning experiment. This approach allows us to study their choices from a continuous set of possible search decisions. Using decision theory we derived the ideal search strategy for our search task. We show that this strategy predicts human search behavior. Our results suggest that people are efficient searchers and that they combine exploration and exploitation strategies in a manner that is close to optimal.

Most studies on human reward learning have been limited to search tasks where there is a discrete set of decisions (Cohen et al., 2007; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Ferguson, 1989; Sonnemans, 1998; Weitzman, 1979). For example, optimal decision policies have been worked out for *n*-armed bandit tasks (Gittins, 1979; Krebs et al., 1978). At least one study has looked at how people make the trade-off between exploration and exploitation in these tasks (Daw et al., 2006). There has even been some proposals of the neural mechanisms and possible brain regions involved in such decision processes (Daw et al., 2006; Yu & Dayan, 2005). Here we have presented an experimental approach that allows us to effectively analyze search behavior when there is a continuous set of search decisions. The reduced dimensionality of the experimental task and the adherence to a fixed unimodal reward function for each trial allow a tractable solution to the ideal search policy. Although the task and optimal solution vary considerably from that of previous work, we show that people are efficient searchers even in the more difficult case where search decisions must be drawn from a continuous possible set.

When a sequence of searches is performed, information obtained from each search should inform how to choose subsequent searches. An optimal search strategy should consider possible future searches and their outcomes. The ability to ''look ahead'' when planning searches significantly improves search performance, since it allows the searcher to make decisions that are optimal in the long run and not just in the current state. This is shown in Fig. 4, where the last reward is on average larger in the simulation of the ideal searcher than in the simulation of the optimal exploitive strategy. To compute the ideal search policy, search locations are optimized over all possible future search decisions. This ensures that the policy employs a full look-ahead strategy. A policy that optimizes only the current search decision, without regard to the future, would prescribe a different set of search locations. The similarity between the participants' search behavior and the ideal policy suggests that people employ a look-ahead strategy. The capacity for look-ahead strategies has already been suggested in human visual search tasks (Najemnik & Geisler, 2005).

An agent should explore more when there are many decisions left and people have been shown to adhere to this principle (Carstensen, Derek, & Charles, 1999). Given that the reward in our experiment depends on the highest score from all searches, the optimal search strategy for the third search should be highly exploratory. The ideal policy prescribes such behavior. The optimal third search position is very close to a pure exploration strategy. At earlier searches or when the number of searches to be performed is unknown, the optimal exploratory search location approximates the golden section of the search space (Kiefer, 1953). Participants' average third search location is at .67, which is close to the golden section (i.e., .62). It would be interesting in future experiments to test how sensitive peoples' strategies are to the manner in which they are rewarded, for example, whether search strategies become more exploitive when rewarded with the sum or the average of reward amounts obtained.

Given the ecological importance of search, we should expect that people are good at searching and behave close to optimal in such tasks. Here we have shown that people can skillfully search in a single dimension when search locations are drawn from a continuous possible set. Participants tend to search close to optimal locations given past rewards and efficiently accrue rewards. People not only demonstrate efficient reinforcement learning strategies when given a discrete set of choices as had been shown previously but also in the case of continuous choice in the motor domain.

References

Bartumeus, F., & Levin, S. A. (2008). Fractal reorientation clocks: Linking animal behavior to statistical patterns of search. *PNAS*, *105* (49), 19072–19077.

Carstensen, L. L., Derek, I. M., & Charles, S. T. (1999). Taking time seriously. *American Psychologist*, *54* (3), 165–181.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society*, *362*, 933–942.

Daw, N., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8* (12), 1704–1711.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decision in humans. *Nature*, *441*, 876–879.

Ferguson, T. S. (1989). Who solved the secretary problem? *Statistical Science*, *4* (3), 282–289.

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience*, *5* (7), 1688–1703.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society*, *41* (2), 148–177.

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*, 780–784.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A Survey. *Journal of Artificial Intelligence Research*, *4*, 237–287.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47* (2), 263–292.

Kiefer, J. (1953). Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, *4* (3), 502–506.

Kohn, M. G., & Shavell, S. (1974). The theory of search. *Journal of Economic Theory*, *9*, 93–123.

Krebs, J. R., Kacelnik, A., & Taylor, P. (1978). Test of optimal sampling by foraging great tits. *Nature*, *275* (7), 27–31.

Maloney, L. T., Trommershäuser, J., & Landy, M. S. (2006). *Questions without words: A comparison between decision making under risk and movement planning under risk*. New York: Oxford University Press.

Mayntz, D., Raubenheimer, D., Salomon, M., Toft, S., & Simpson, S. J. (2005). Nutrient-specific foraging in invertebrate predators. *Science*, *307*, 111–112.

Montague, P. R., Dayan, P., Person, C., & Sejnowski, T. (1995). Bee foraging in uncertain environments using predictive hebbian learning. *Nature*, *377*, 725–728.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*, 387–391.

Pyke, G. H. (1984). Optimal foraging theory: A critical review. *Annual Review of Ecology and Systematics*, *15*, 523–575.

Sonnemans, J. (1998). Strategies of search. *Journal of Economic Behavior & Organization*, *35*, 309–332.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*, 1782–1787.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.

Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor control. *Nature Neuroscience*, *5* (11), 1226–1235.

Trommershauser, J., Maloney, L., & Landy, M. (2003). Statistical decision theory and the selection of rapid, goal-directed movements. *Journal of the Optical Society of America A-Optic Image Science and Vision*, *20* (7), 1419–1433.

Viswanathan, G. M., Buldyrev, S., Havlin, S., da Luz, M. G. E., Raposo, E. P., & Stanley, H. E. (1999). Optimizing the success of random searches. *Nature*, *401*, 911–914.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica*, *47* (3), 641–654.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*, 681–692.