Signatures of visuospatial information foraging during learning of complex environments

David L Barack[1], Akram Bakkour[4], Daphna Shohamy[2], C Daniel Salzman[1,3]
[1]Department of Neuroscience, [2]Department of Psychology, and [3]Department of Psychiatry, Columbia University; [4]Department of Psychology, University of Chicago

Abstract

Many real world environments contain latent features that animals must learn in order to accomplish their goals. Animals often learn these environments over sequences of choices, but the behavioral mechanisms underlying this complex cognitive competence remain poorly characterized. We investigate this sophisticated behavior in two primate species, humans and macaque monkeys, with a task in which subjects searched for shapes hidden on a grid. Both primate species were adept learners, quickly learning the subset of shapes out of the numerous possibilities. Subjects tended to select tiles earlier in trials that were informative in the past about the hidden shape than those that were rewarding. In addition, we found a surprising signature of foraging behavior over sequences of choices during trials, with human subjects searching local areas of the board until information dropped below the average across all choices, at which time they jumped to a different part of the board. This pattern of choices was not evident for rewards in humans. In contrast, the sequences of choices of monkeys were equally well-described as information or reward foraging. Finally, the rate at which humans learned shapes could be predicted by how well their choice sequences matched foraging behavior. These findings suggest that humans are more tuned to the search for information than reward than monkeys and that foraging competence predicts the capacity to learn complex environments.

Introduction

Most animals must learn about the world around them, the features of their environments and the causal relations between them [1-7]. Real world environments possess innumerable features and causal relations. This complexity can be conceptualized in terms of the number of states the environment contains. The features and causal relations that occur together constitute the states of the environment. Simple environments may contain only a few states, whereas complex ones may contain thousands or more. Despite the evident complexity of the real world, most tasks in psychology, neuroscience, and other disciplines are relatively simple and contain few states.

How can animals learn such complex environments? Animals could use rewards or information to learn environments. Features of the environment often correlate with rewards, thereby providing animals with feedback that can be used to augment behavior [8-10]. However, the complexity of the environment or the number of choices required to gain rewards present challenges for this use of rewards. Information, here understood as changes in the distribution of features or causes that characterize environments, can also be used to learn. Animals can learn their environments by making observations about what features occur together or what features

seem to stand in causal relations [3, 6, 8, 11-21]. These changes need not rely on received rewards and so may present a more fruitful strategy for learning complex environments.

Recent research has probed how animals learn their environments with behavior [3, 22-25], neuroimaging [26-29], or electrophysiology [30], or approached the topic from the perspective of inferences about the causal relations between environmental features [2, 3, 18, 29, 31-43]. Many of these tasks rely on features that are perceptible. While perception can help learn environments based on observable properties, features of environments that are completely or partly unobservable (based on so-called latent states) must be inferred [44, 45]. Most real world complex environments contain latent features that manifest in many perceptual and causal ways [46]. In addition, many of these tasks probe how environments are learned with single choices on each trial. However, often the environment is progressively revealed over multiple choices. Our research aimed at uncovering the behavioral mechanisms that allow learning over multiple choices in such complex environments.

To assess learning with sequences of choices in a complex environment, we developed a novel behavioral paradigm based on the board game Battleship that can be played by both humans and monkeys. On each trial, subjects searched through and selected squares from a grid to reveal whether or not a piece of a shape is hidden beneath the square. Subjects made sequences of such choices to finish revealing shapes. Our task is a multi-alternative sequential search task where during trials subjects learn to reveal the current shape and learn over trials which shapes out of thousands of possibilities actually occur in the environment. We had both human and nonhuman primate subjects perform our task. Subjects from both species were very adept at the task, quickly learning where shapes lie on the board. The analysis of the behavior revealed that information seeking and not reward seeking drove choices. Specifically, tiles that were historically more informative tended to be selected sooner in trials, whereas historically more rewarding tiles tended to be selected later. Importantly, the pattern in which subjects of both species sampled the board correlated with momentary decreases in information intake below the average for the environment, a type of foraging behavior observed in many other contexts across species. In sum, evidence from choice behavior on our task suggests both humans and monkeys relied on information search to learn shapes and foraging computations to decide where to sample spatially on the board to gather information. This exquisite skill for information search across different species suggests an adaptive specialization of the primate clade for efficient information-driven learning of the environment.

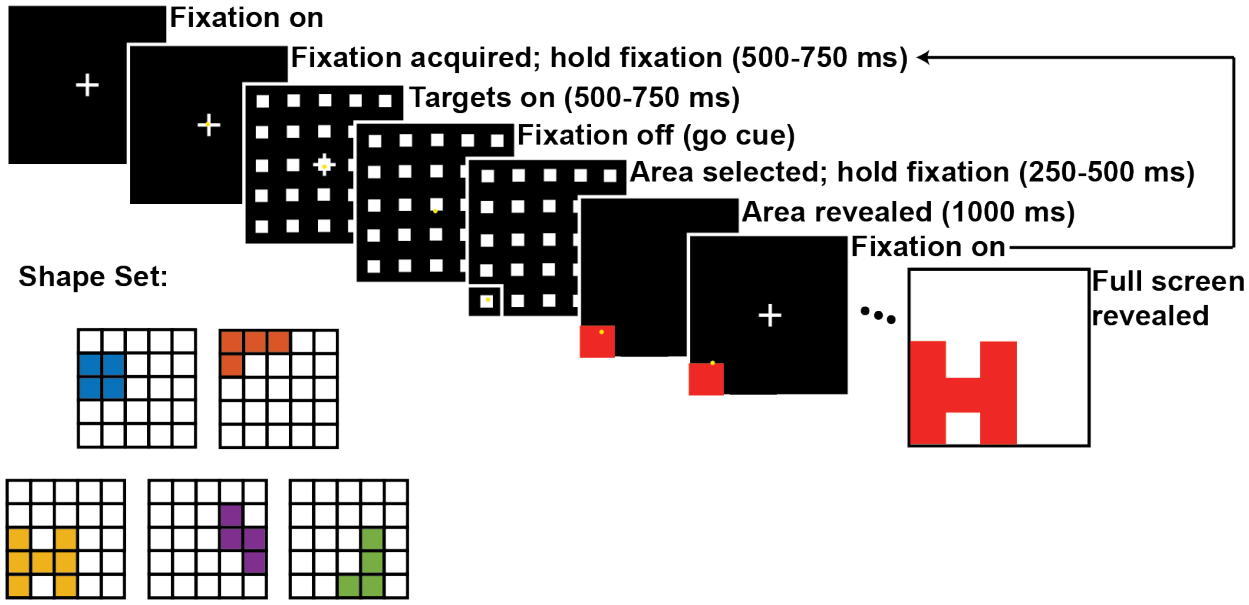
Results

To study how animals learn complex environments with thousands of possible states, sets of features that occur together, we designed a shape search task (the ‘battleship’ task). In this task, subjects viewed a 5x5 tile grid and chose locations on the grid to reveal either a filled or empty square. In the following, ‘board’ refers to the set of possible configurations of filled, empty, or yet-to-be-selected tiles, ‘state’ refers to the set of possible configurations of filled or empty tiles on fully revealed boards, and ‘shape’ refers to one of the set of states that the animal could actually

experience. On every trial, one of 5 possible shapes was initially hidden on the board. Subjects make choices to reveal the hidden shape by selecting tiles (Fig. 1A). The 5 possible shapes that can be revealed are only a small subset of the 3,904 possible states in the task (see methods). The set of five shapes did not change during the task and each shape always occurred at a given location on the game board, though subjects were not instructed with regard to either the number of shapes or locations. Shape colors and identities varied from trial to trial.

After a choice, if subjects uncovered a part of a shape ('hit'), then they received a reward (points for humans, juice for monkeys). If subjects failed to uncover a part of a shape ('miss'), then they proceeded to the next choice. Every choice registers as a hit or a miss. Trials were complete once all filled squares in the shapes were uncovered. At trial start, subjects made a movement to a target at the center of the screen (humans: mouse-over; monkeys: saccade), maintaining position on the fixation point while targets appeared in the middle of each remaining tile. After a variable delay, the fixation point disappeared and subjects had unlimited time to choose a target. A hit or miss was revealed at the chosen tile location. After an inter-choice interval, the fixation point reappeared. Subjects had to reacquire fixation between every choice. This sequence repeated until the shape was fully revealed, which was followed by a longer free viewing period and then an inter-trial interval (see methods).

A


Fixation on
Fixation acquired; hold fixation (500-750 ms)
Targets on (500-750 ms)
Fixation off (go cue)
Area selected; hold fixation (250-500 ms)
Area revealed (1000 ms)
Fixation on
Full screen revealed
Shape Set:

B

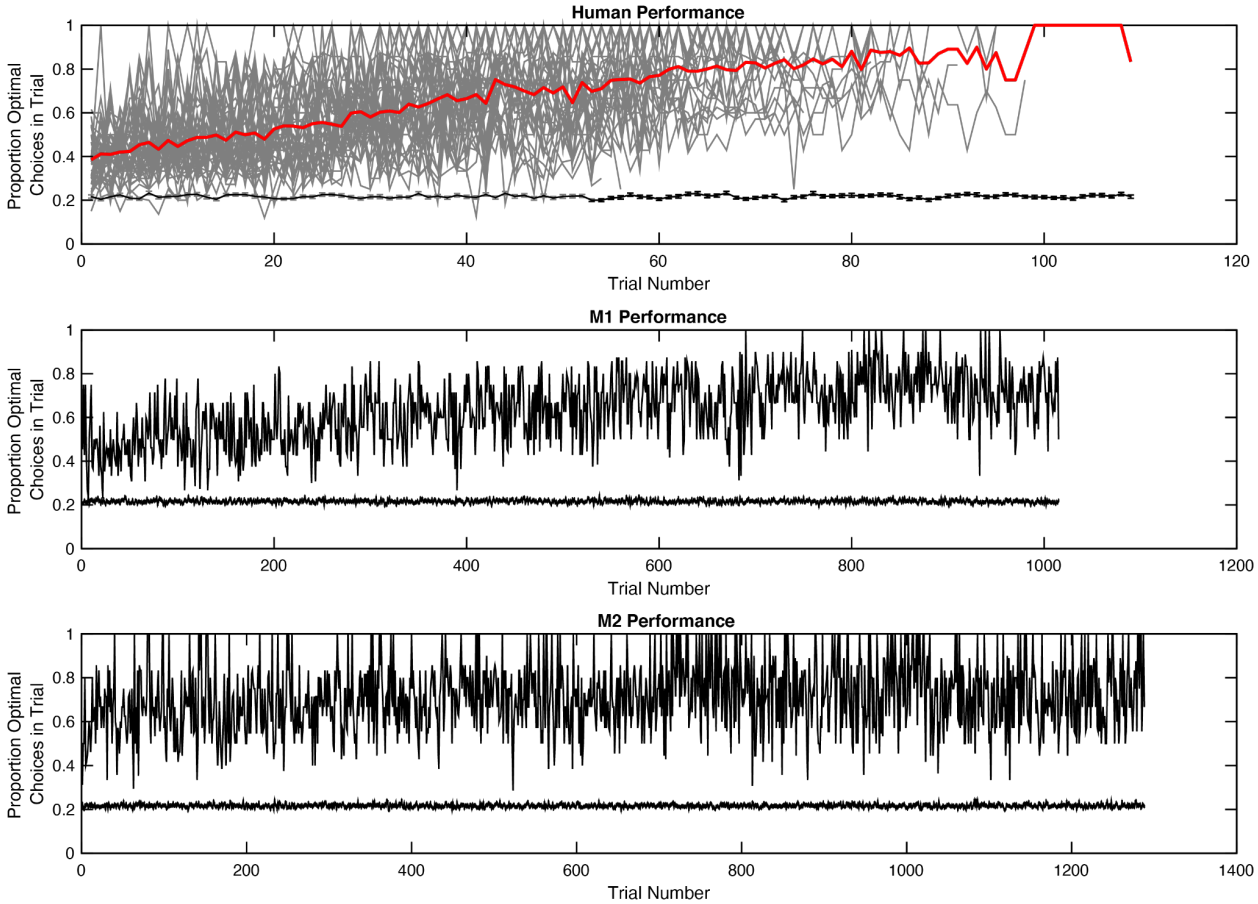

Human Performance
M1 Performance
M2 Performance
Proportion Optimal Choices in Trial
Trial Number

Figure 1. A. *Battleship task. The variable time intervals reported in the figure are for the monkeys; the human subjects had hold times of 500 ms for fixation, 500 ms to after targets on, and 250 ms to register selection of a target.* ***B.*** *Performance on the battleship task, x-axis: trial number, y-axis: proportion of optimal choices in trial. Top panel: thick red line = average human performance (n = 42), thin gray lines = individual subject traces; middle panel: M1; bottom panel: M2. For all three panels, jagged line at bottom is the average simulated performance of a choice algorithm that randomly chooses tiles (1000 iterations).*

To assess performance, we plotted the proportion of optimal choices by trial (humans: 45 min session; M1 and M2: all runs concatenated across 7 days). An optimal choice was defined as the choice that maximized expected reward given what had been revealed on the board so far. To maximize expected reward, the subject should choose the tiles where the remaining shapes have the most overlap. The proportion of optimal choices increased over learning for both species, with human subjects an order of magnitude faster than monkeys in stabilizing the proportion of optimal choices per trial around 70% (Fig. 1B). All subjects outperformed an algorithm that randomly selected tiles on every choice (Fig. 1B, small dashed lines below data).

Granted both humans and monkeys learned to reveal the shapes, did rewards or information drive adjustments in subjects' behavior? Past outcomes on the battleship task deliver both rewards (points for humans and juice drops for monkeys) and information (changes in the probabilities of a pattern of filled and empty squares on the fully revealed board). Information is defined in terms of changes in the entropy of the probability distribution (see methods). The entropy of the distribution is computed before the outcome and after the outcome. The information gain is equal to the difference between these two entropies. We use a Dirichlet distribution to model the distribution of probabilities of the different states across trials (see methods). Informally, this distribution can be thought of as a set of categories, one for each state, that keeps track of how many times each state has been observed. The probability of any given state is determined by the frequency with which that state is experienced. The initial counts for each state at the start of each trial were the updated counts from the end of the previous trial. These counts were transformed into probabilities by dividing the count for each state by the total count across all states.

During trials, a different distribution '$\mathfrak{B}$' is initially set to the values in the Dirichlet distribution. After the outcome from each choice in the trial, $\mathfrak{B}$ is updated using a Bernoulli vector (i.e., a vector of 1's and 0's) for the likelihoods of each state given the outcome of that choice. For a given state, if that state was consistent with the outcome, then the likelihood was 1; otherwise the likelihood was 0. The prior probability was multiplied by the likelihood of the outcome to yield a posterior for each state. Each state's probability was then divided by the sum of the probabilities across all states to attain the new $\mathfrak{B}$, the prior for the next choice in the trial. At the end of the trial, the Dirichlet distribution was updated by adding a single count to the bin corresponding to that trial's state (see methods).

In order to assess the influence of information and reward on choice, a multinomial logistic regression was performed to regress choice number in trial against trial number in run, the

summed history of information for the chosen tile, and the summed history of rewards for the chosen tile (Fig. 2). For human subjects, the mean coefficient of summed information was positive and significantly different from the negative mean coefficient for summed reward ($t(df=670) = 17.2822$, $p < 5x10^{-55}$; mean $\beta_{info} = 3.02 \pm 1.51$; mean $\beta_{reward} = -1.62 \pm 0.86$). In humans, summed information correlated with higher probability of earlier choices of tiles whereas summed reward correlated with higher probability of later choices of tiles. Paired t-tests across choice numbers revealed that information influenced choice significantly more than reward ($t(df=14) = 6.6300$, $p < 5x10^{-5}$), confirming this finding.

The same outcomes correlated with the choices of monkeys. As in humans, information more positively influenced choice than reward in both monkeys, although M1 was more driven by information than M2 ($t(df=7) = 4.1530$, $p < 0.005$) whereas M2 was more driven by reward than M1 ($t(df=7) = -3.1145$, $p < 0.05$). The influence of information on choice was greater than reward in M1 ($t(df=7) = 6.1807$, $p < 0.0005$) but not M2 ($t(df=7) = 1.3794$, $p > 0.2$). This inter-subject variability may provide leverage for understanding neural mechanisms of such choice.

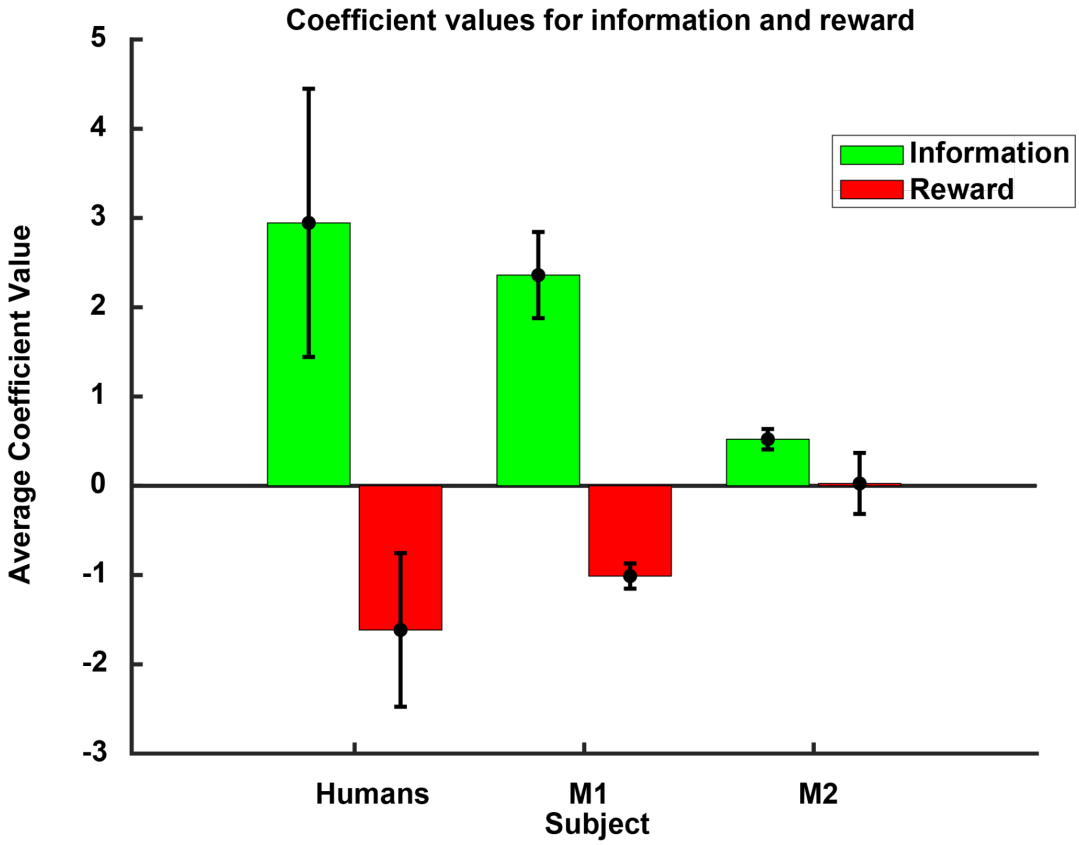
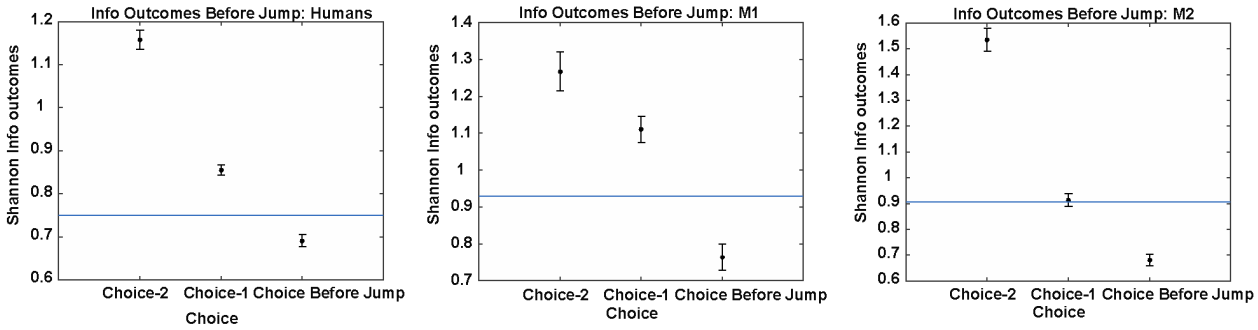


Figure 2. *The influence of information and reward on choice number in trial for humans, M1, and M2. Each column is the average beta (left pair of bars, across human subjects and choice numbers, mean by subject ± 1 s.e.m.; middle and right pairs of bars, M1 and M2 respectively, across choice numbers, mean by choice number ± 1 s.e.m.).*

The average coefficients for human subjects for each choice number were compared to each monkey using paired t-tests. For information, the influence of summed information on humans was greater than on M1 ($t(df=7) = 4.5787$, $p < 0.005$) or on M2 ($t(df=7) = 4.4092$, $p < 0.005$). For reward, the influence of summed rewards on humans was marginally more negative than on M1 ($t(df=7) = -1.9364$, $p < 0.1$) and significantly more negative than on M2 ($t(df=7) = -2.5719$, $p < 0.05$). Finally, the relative influence of information over reward in humans was significantly different from M1 ($t(df=7) = 2.8571$, $p < 0.05$) and M2 ($t(df=7) = 3.4454$, $p < 0.05$). In sum, information encouraged earlier choices than reward in humans and monkeys, though the difference in this influence was greater in humans than in monkeys.

Besides single choices, information or reward may predict aspects of sequences of choices as well. Let a jump be defined as a choice of a tile more than one square away from the previous choice (i.e., a choice of a non-neighboring tile). The information and reward outcomes from sequences of three choices of neighboring squares in a row before a jump were plotted. A significant number of trials contains such sequences (humans: 0.28 ± 0.01 fraction of trials; M1: 0.27; M2: 0.29). These plots revealed that both human and monkey subjects tended to jump after these sequences of choices when the information intake dropped below the average information intake across all choices (Fig. 3A), a pattern diagnostic of foraging behavior [47]. In humans, this pattern was not observed when these choices were plotted using reward outcomes (Fig. 3B, left panel). In contrast, monkeys' jumps following outcomes were equally well-described by falling below either average information or reward outcomes (Fig. 3B, middle and right panels). In sum, human choice sequences were better described as foraging for information than reward, whereas monkey choice sequences were equally well described as foraging for either resource.

A



B

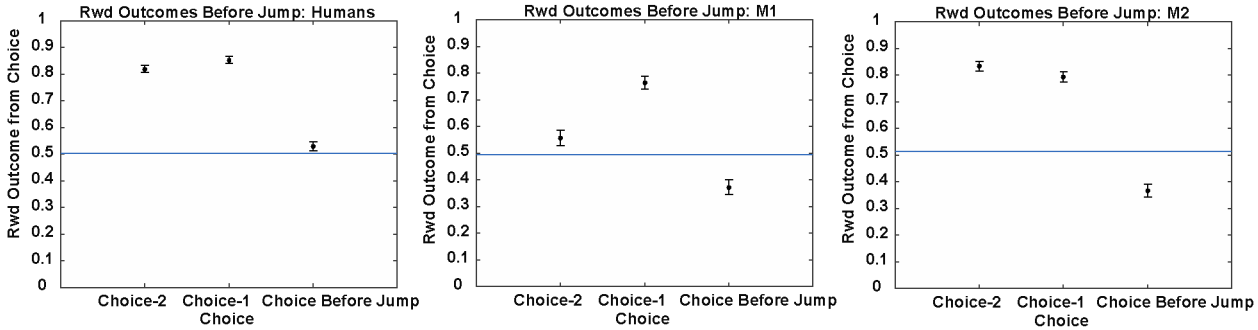


Figure 3. A. *Left: human information foraging; Middle: M1; Right: M2.* ***B.*** *Left: human reward foraging; Middle: M1; Right: M2. Note that the reward outcome for a choice before a jump for humans was not below the average reward outcome across all choices. Both* ***A*** *and* ***B*** *include all trials during and after learning.*

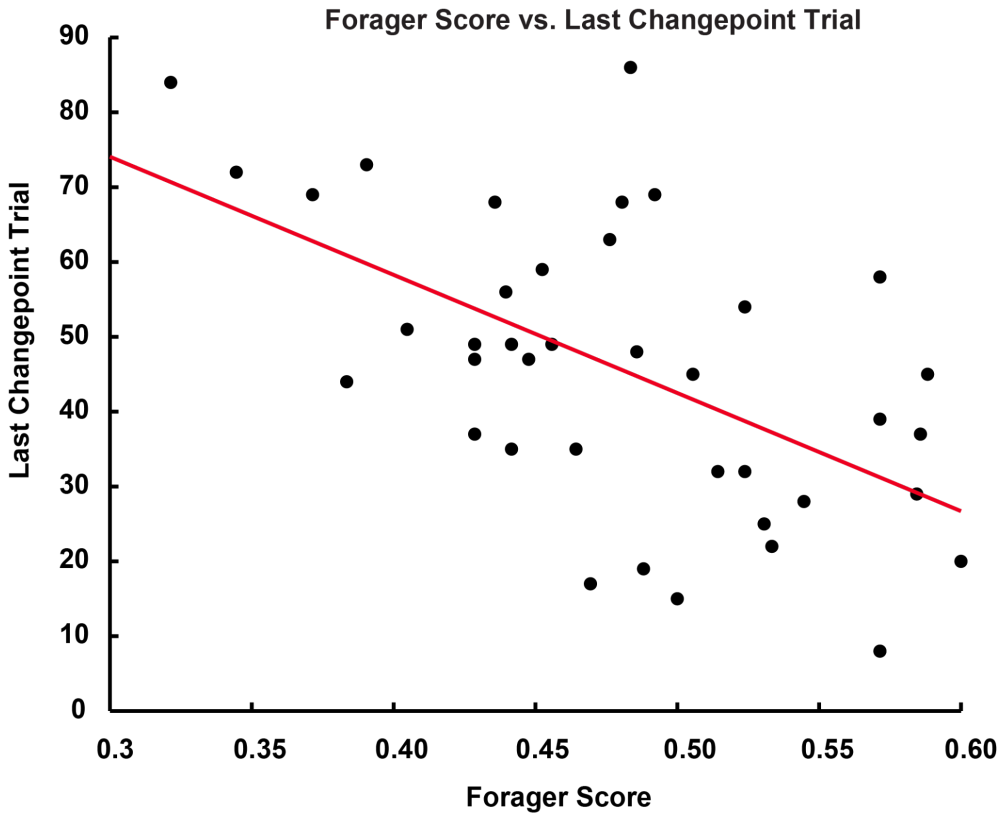


Figure 4. *For human subjects, the forager score F_s predicts the speed of learning. Each point is a single subject. Red line: OLS regression (β_{slope} = -112.75 ± 27.08 trials until changepoint/a.u. forager score, $p < 0.0005$).*

A quantitative description of choice behavior in foraging terms correlated with faster learning. To quantify forager performance, a forager score was computed for each subject during and after learning (see methods). A method to determine the end of learning is needed to quantify the speed of learning. To determine the end of learning, we reasoned that two main effects of learning on choices should be evident. First, the mean number of choices to finish revealing a shape should diminish during learning. Second, the variance in the number of choices should diminish as subjects learned the shape or movement representations [cf. 48]. To detect changes in the mean and variance, a changepoint detection test was run on the mean and variance of subjects' choices across all shapes and trials (see methods) [49, 50]. The end of learning was set to the last changepoint (whether mean or variance) across all shapes.

Better information foraging scores for humans during learning predicted earlier final changepoints and hence faster learning (Fig. 4; OLS, β_{slope} = -112.75 ± 27.08 trials until changepoint/a.u. forager score, $p < 0.0005$, $\rho = -0.56$). No such relationship was found for reward foraging scores (OLS, β_{slope} = 41.84 ± 43.99, $p > 0.3$, $\rho = 0.15$). During learning, human forager scores (mean_forager_score = 0.67 ± 0.02) were not significantly different from M1 (M1 = 0.67; t(df=38) = 0.3596, $p > 0.7$, one sample t-test) and only marginally different from M2 (M2 = 0.70; t(df=38) = -1.9030, $p > 0.05$, one sample t-test). After learning, humans (mean_forager_score = 0.56 ± 0.04) had lower information foraging scores than both monkeys (M1 = 0.72, t(df=37) = -3.8197, p <

$5x10^{-4}$; M2 = 0.72, t(df=37) = -3.8413, $p < 5x10^{-4}$; both one sample t-tests). Further, humans foraged significantly more during learning compared to after (paired t-test, t(df=75) = 2.6313, $p < 0.05$). In sum, better foraging in humans predicted faster learning (as judged by their last changepoint) and, unlike monkeys, humans ceased information foraging after learning.

Discussion

Learning complex environments with thousands of states is a difficult challenge for any organism. We implemented a shape search task on a tiled grid with thousands of possible states to investigate how human and nonhuman primates learn complex environments. We discovered that both human and nonhuman primate subjects preferred to choose tiles sooner in trials that were historically informative in comparison to reward and that drops in information outcomes predicted decisions to sample new areas of the board.

Strategies to learn simple or complex environments can be motivated by the search for reward or for information. Common reward-based approaches such as classic [51] and computational [8] reinforcement learning assume a state space. Rewards can be used to learn environments by assigning credit for them to co-occurring features, that is, states of the environment [52]. However, complex environments with many states require long sampling histories for learning with such reinforcement processes [53], although numerous improvements have been suggested for these computations, such as the introduction of options that group together many actions [54]. In a relatively sparse environment where most states do not deliver rewards, animals that learn using only rewards may require many trials over long periods of time. Instead, information in the form of changes in the distribution of environmental features can be used to speed up environmental learning. Such information-based approaches that explore options to gather information to subsequently maximize rewards can be more efficient [3, 6, 8, 11-21, 55, 56]. These processes aim to reduce uncertainty, formalizable in terms of information theory [57, 58], about the environment by determining which features tend to occur together.

Recent research has begun to elucidate how primates search for and use information instead of reward to learn their environments. Since animals often do not know the most informative choices, they must rely instead on hypothesis testing[59-68], information foraging[69-75], or maximizing expected information gain[20, 21, 76-80]. Humans are excellent information seekers [81], efficiently using information to learn their environments [20, 21] and reason about the world in pursuit of their goals [76, 82-86]. Recent research in human information seeking suggests that humans sometimes select options that maximize expected information gain from outcomes {Coenen 2018}. This tendency occurs in reasoning tasks like the Wason selection task [64, 76, 87], categorization tasks [84, 85], and many others (see [88] for review). Two aspects of the real world are often absent in such tasks. First, many of these tasks probe subjects' preferences for information with a single choice. The grounds of sequential information search, in contrast, are just beginning to be explored [89, 90]. Second, many of these tasks involve few states. This is true of most single choice tasks [see, e.g., 76] but also of those sequential search tasks [such as 86, 89, 90, 91, 92].

Besides humans, the extent to which other animals search for information and how efficiently is less well characterized. Many animals show a preference for advanced information about outcomes that cannot be used to adapt behavior (so-called 'observing responses'; [93]), including pigeons [94], monkeys [95, 96], and humans [97, 98]. These studies tend to use environments with very few states. In addition, animals' preference for usable information remains poorly characterized. Though monkeys in particular show these preferences [99, 100], the way and degree to which monkeys search out usable information is just beginning to be described. More generally, the presence of a competence for information search across species would suggest broad selective pressures for such behavior.

We uncovered general preferences for information across species. Individual choices were driven by information in comparison to reward. Specifically, tiles that were more informative in the past tended to be selected sooner in trials than historically rewarding ones. Human subjects selected historically informative options sooner than historically rewarding ones. The evidence for monkeys was split. One monkey significantly selected informative tiles sooner than rewarding ones, while another monkey did not exhibit a significant difference in the impact of rewards and information on choice, though the same basic pattern of preference for information over reward was present in both.

This use of information also correlated with features of sequences of human subjects' choices. A significant number of subjects' trials across species showed sequences of choices that reflected an area restricted search, persistent searching through a limited spatiotemporal region. Area restricted search is a hallmark of foraging behavior [47, 101-104]. A standard computation during foraging is to use the average reward in the environment to guide decisions to continue foraging locally or to leave the local area to look for new resources [105]. Once intake rates drop below the average for the environment, subjects should depart the local area and search for new resources.

We observed signatures of such information foraging behavior while subjects searched for parts of shapes. During learning, both humans and monkeys engaged in area restricted search for information. Further, decisions to search a different part of the board were predicted by drops in information intake below the average across all choice outcomes, a signature of the average threshold rule for decisions to leave the local area of the board. A similar pattern was observed for rewards in the monkeys but not in the humans.

Information foraging is a key topic for understanding cognition [69, 71, 102, 106]. Information foraging has been previously observed for humans surfing the internet [70, 71, 107, 108]. In one study [70], human subjects chose to click to leave a webpage when the information remaining on the page as determined by an 'InfoScent' model dropped below the average across all pages. A simple model-free information foraging behavior has not before been observed, however. The search for information forms the basis for understanding visual search [72, 109, 110] or chemotaxis[73]. The foraging framework generally has also been extended to numerous cognitive processes, including task-switching [111], internal word search [112], study time allocation [113],

problem solving [114], memory search [115], semantic search[116], and even social interactions [117]. The foraging effects in these studies typically regard reward-based decisions to shift between options or decrements in behavior that induce shifts between options. Information search also plays a role in explaining behavior on numerous tasks, including past studies on battleship-like search tasks [20, 21, 118]. These studies do not report evidence for an information-intake threshold rule such as reported here.

While we report evidence for visuospatial information foraging in a complex task that predicts learning for the first time, our results have several key limitations. First and foremost, we used a small number of shapes and tested a single shape set. While the shape set was selected because of the opportunity to compare learning in human to nonhuman primates, these findings remain to be generalized to other shape sets and other contexts such as bigger boards. Second, our results crucially rely on various assumptions that can be challenged. For example, we used inferred changepoints as a measure of learning. But changepoints in the mean or variance of choices might instead be the result of other processes such as attention, arousal, or boredom. We also assumed that subjects' choice behavior can be understood as foraging in a restricted area, some contiguous set of squares on the board, and decisions to choose a non-neighboring tile reflect decisions to move to a different restricted area. This approximation to 'patches' from foraging theory is ad hoc. Future experiments will need to manipulate confounding factors like attention or test assumptions about restricted areas to ensure our results don't reflect such assumptions. Third, our study focuses on information in light of the pursuit of reward. While our findings suggest that information foraging can result in faster learning and higher long-term reward rates, our study does not probe the search for information for information's sake [119]. Such intrinsically motivated information search remains difficult to probe, especially in nonhuman animals that require special alimentary motivation. Consequently, our results may not stand in contexts where the search for information is its own reward. Fourth, we did not model the choice algorithms underlying decisions on the task. Our results, however, can be used to construct such algorithms and we intend to use our findings as a guide to future modeling.

Conclusion

While many studies have investigated how animals learn their environments, few have explored this learning in complex environments with thousands of states using sequences of choices. We discovered that in a complex sequential decision making task with thousands of possible shapes, humans and monkeys were both driven more by information than reward and foraged for information by sampling different areas of the board in line with predictions from foraging theory. This fundamental computation suggests that decision making circuitry evolved for searching for nutrients and other environmental resources is also used to learn about the environment using more abstract resources like information. Additionally, strategic adjustments in foraging behavior observed in humans but not monkeys suggests that hominids may have more exquisite adaptations for searching for and using information.

Methods

Our battleship task requires subjects to uncover shapes by selecting tiles that hide parts of shapes. Herein, 'hits' refers to choices that revealed part of a shape and were rewarded, and 'misses' refers to choices that did not reveal part of a shape and were not rewarded. Rewards were points for humans and squirts of juice for monkeys. We used five distinct shapes composed of multiple tiles on a 5x5 grid (Fig. 1A; 1 shape ('H') occupied 7 tiles, and the other occupied 4 tiles). A trial refers to the sequences of choices required to finish revealing the shape. There was one shape to uncover per trial, pseudorandomly drawn from the set of five. The color of shapes pseudorandomly varied from trial to trial. Shapes occurred at the same location across trials and subjects, with a different location for each, and shapes overlapped at certain tiles. The shapes were selected on the basis of pre-existing data for both monkeys that was collected while training them for a distinct electrophysiology study and were chosen by the experimenter (DLB).

At trial start, subjects made a movement to a target at the center of the screen (humans: mouse-over; monkeys: saccade), maintaining position on the fixation point (humans: 500 ms; monkeys: 500-750 ms) after which targets appeared in the middle of each remaining tile. After a second delay (humans: 500 ms; monkeys: 500-750 ms), the fixation point disappeared and subjects had unlimited time to choose a target. To select a tile, subjects made a movement (humans: mouse-over; monkeys: saccade) to a target at the center of the tile and held their position (humans: 250 ms; monkeys: 250-500 ms). A hit or miss was then revealed at the chosen tile location. After an inter-choice interval of 1 sec, the fixation point reappeared. Subjects had to reacquire fixation between every choice. This sequence repeated until the shape was fully revealed, which was followed by a 2 sec free viewing period and then a 1 sec inter-trial interval.

Monkeys (*M. mulatta*) performed the task sitting in a primate chair (Crist) with their heads immobilized using a custom implant while eye movements were tracked (EyeLink; SR Research). All surgeries to implant head restraint devices were performed under strict IACUC approved protocols (Columbia University) in fully sterile surgical settings. After recovery from surgery, animals were first trained to look at targets on a computer screen for squirts of juice. They were then trained to make delayed saccades, maintaining fixation on a centrally presented square while a target in the periphery appears. Once the central square extinguished, animals could then make an eye movement to the target. Next they were trained on the battleship task, starting with a 3x3 grid, then a 3x4, and so on until a 5x5 grid. The data presented herein are from the first set of shapes both monkeys learned on a 5x5 grid.

Humans (*H. sapiens*) performed the task using a mouse on a computer. The task was programmed in javascript. They received the following instructions before the first trial:

> *Hello! Welcome to the battleship task!*
>
> ---
>
> *The goal of this task is to identify the shapes*
> *with the fewest number of searches. To select a shape*
> *please hover over the central fixation point for 0.5 seconds*
> *After the targets appear, remain fixed on the central point for*

another 0.5 seconds until the blue central fixation point disappears. Then you are free to select a target by hovering over the target in the desired square. Once you have found a shape, the task will start over. Thank you, and you will be debriefed at the end. Good luck!

As indicated by the instructions, human subjects chose tiles by mousing over the target until the choice registered and that part of the board revealed. There were no training trials and we imposed no exclusion criteria on subjects for number of trials for analysis. We collected 42 human subjects (16m, 26f, average age 22 ± 4.9), recruited from the New York City community around Columbia University (most were Columbia University students). All procedures were approved by the Columbia University IRB.

Performance on the task was assessed by calculating the proportion of optimal choices on each trial. An optimal choice is a choice of a tile that maximized expected reward given what has been revealed before that choice. For example, suppose that no tiles have been chosen yet at the start of the trial. Then, the optimal choice is to select one of the four tiles at which two shapes overlap, because given what has been revealed so far (i.e., nothing), those tiles maximize the expected reward. For a second example, suppose that the first choice in a trial eliminates all but two of the shapes. And suppose that the two remaining shapes do not overlap at any further tiles. Then, the choice of any tile that is part of either of the two shapes, but no other tiles, would be optimal because given what has been revealed so far, only those tiles have any expected reward and they all have the same expected reward. For each choice in each trial, the expected rewards for each remaining tile were calculated, and then the subjects' actual responses compared to these expectations. If the subject chose any one of the tiles with the highest expected reward, then the choice was considered optimal; if not, then the choice was suboptimal. The proportion of these optimal choices was then computed for every trial and plotted in Fig. 1B.

To compare this behavior to some baseline choice strategy, we simulated an agent using a random choice strategy: select any one of the remaining tiles with equal probability after each outcome. For humans, simulations were performed for the maximum number of trials across all subjects (Fig. 1B, top panel), and for monkeys, for the same number of trials as each subject (Fig. 1B, middle and bottom panels). For each simulated trial, a shape was randomly drawn from the set of five, and the simulated random chooser selected tiles until the shape was completed. The simulation was iterated 100 times, the optimal performance computed for each trial, and then these optimal performances were averaged and the standard error of the mean (s.e.m.) computed and plotted (Fig. 1B).

To quantify how quickly subjects learned, we used a changepoint detection test on the mean and variance of choices for each separately [49, 50]. The changepoint detection test (cf. Gallistel et al. 2001) takes the cumulative sum of the number of choices used to finish a shape and looks for changes in the rate of accumulation. During learning, subjects would take many choices to finish revealing a shape. The cumulative sum over these trials would rise correspondingly quickly. Once shapes were learned, however, the cumulative sum would rise more slowly as fewer choices are required to finish revealing a shape. At each successive trial, this cumulative sum is calculated, and the log-odds of a change in slope are computed and tested against a changepoint detection

threshold (set to 4, corresponding to $p < 0.001$). The changepoint detection test was run on the set of trials for each shape separately. In addition to changes in the mean number of trials to finish revealing a shape, a change in the variance of the number of choices over some window may also signal learning; as subjects learn, they will become less variable in the number of choices needed to finish a shape. After running the changepoint detection test on the cumulative sum of choices, the value of the best-fit line through each detected changepoint interval was subtracted from the number of choices on each trial. The result is a vector of residuals for the number of choices to finish revealing a shape. The variance over a moving window of 5 trials was then computed for these vectors, and the changepoint detection test performed on the cumulative sum of that variance [cf. 50]. Monkeys but not humans learned the shapes over multiple days. This variance was not calculated for sets of 5 trials that spanned days. However, changes in biological and psychological processes irrelevant to learning, such as arousal, wakefulness, and so forth, can spuriously contribute to the calculated variances. A day-to-day variance correction was performed to control for this: the variance in the 5-trial-wide window was divided by the global variance across all trials and shapes for the respective day. The cumulative variance changepoint detection test was then performed on the variance-normalized-by-day data. The end of learning was set to the last detected changepoint trial across both mean and variance changepoint detection tests. This test failed to detect a changepoint for 3 of 42 human subjects, who were removed from the learning analyses as a result.

A distribution of probabilities for each possible pattern that the board could take once fully revealed was used to compute information outcomes. Here, 'state' refers to these possible fully revealed boards. These boards were formed by possible combinations of connected, filled tiles on a 3x3 grid that were then placed on the 5x5 board. The shapes had to be at least 3 tiles large and connections had to be in the vertical or horizontal directions (i.e., no diagonal-only connections permitted). To simplify information calculations, the set of states ($n_{state} = 3904$) was assumed to be known to the subjects. While both monkeys had been trained on smaller boards using these possible states, the humans were naïve to the task and to the distribution of states. Consequently, this assumption is strictly speaking false for humans, but we adopt it for numerical reasons.

A Dirichlet distribution, which possesses a conjugate prior when using a multinomial likelihood, modeled the probabilities of the states across trials. The Dirichlet distribution sits on the K-1-simplex such that it can be conceptualized as a distribution of K-dimensional distributions. The dimensionality K in the battleship task refers to the number of possible states.

The distribution of the probabilities of the states, which we label '$\mathfrak{B}$', was updated during trials as follows. At the start of the trial, $\mathfrak{B}$ was set to the values in the Dirichlet distribution. After each choice in the trial, $\mathfrak{B}$ was updated using a vector of 1's and 0's for the likelihoods of each state given the outcome of that choice. If a given state was consistent with the outcome, then the likelihood was unity; otherwise the likelihood was zero. The prior probability was multiplied by the likelihood of the outcome to yield a posterior for each state. The resulting probabilities were then re-normalized to attain the new $\mathfrak{B}$, the prior for the next choice in the trial. Information was defined as the difference in the Shannon entropy H of $\mathfrak{B}$ before and after a choice outcome:

$$H_{\mathfrak{B}} = -\sum_{i=1}^{3904} p(x_i) * log\ (p(x_i))$$

for probability of state $p(x_i)$. This assessment of information intake intuitively reflects how much a hit or a miss from a choice changes the subject's uncertainty about the current trial's state. At the end of the trial, the Dirichlet distribution was updated by adding a single count to the bin corresponding to that trial's state, and the maximum a posteriori estimate of the distribution recalculated.

To ensure numerical stability and proper updating using the maximum a posteriori estimate of the Dirichlet distribution, some initial number of samples for each state is required. The Dirichlet distribution is characterized in part by the use of so-called inertial priors described by a series of parameters $\alpha_i, \ldots, \alpha_n$ in a multivariate Beta distribution. These priors are akin to the assumption that each state has been sampled some α_i number of times. Here we assume all the α_i are equal. For each state, the larger these values, the larger the number of new samples needed to shift the probability mass of the distribution away from those states. The maximum a posteriori estimate subtracts one from each bin and conceptually bin counts cannot be less than 0, suggesting a single count for each bin. However, an initial count of 1 would yield a division by 0, so for numerical stability some value above 1 is needed. Since states are either seen or not seen, conceptual considerations suggest an integer value, and 2 was chosen as the simplest, smallest, model-free numerically stable initial count for each state.

In order to investigate the influence of past rewards and past information outcomes on choice, we performed a multinomial regression (*mnrfit* in MATLAB). The dependent variable was the choice number in the trial. The z-scored independent covariates included trial number in run, cumulative sum of past rewards for each tile, and cumulative sum of past information outcomes for each tile. No interactions were included in the regression. The regression fits the following model to the data:

$$\log (p_{i<j} / p_{i \geq j}) = \beta_0 + \beta_1 * t_n + \beta_2 * I_{1:t-1} + \beta_3 * R_{1:t-1}$$

for choice numbers i, j, trial number t_n, summed history of information outcomes $I_{1:t-1}$ from trial 1 to trial t_{n-1}, and summed history of reward outcomes $R_{1:t-1}$ over the same range. In addition, we truncated the data to consider only those choices before the 10th choice in a run; this truncation was performed in order to exclude choices that were very late in trials due to inattention, fatigue, or indolent choice strategies and that had very few choice counts, as well as to focus on those tile choices that were motivated, as opposed to forced as a result of fewer options on the board. The results of this regression are plotted in Figure 2.

To plot sequences of choices, each choice on each trial was sorted according to whether or not the previous choice had been a neighboring tile. Those choices that were non-neighboring were labeled 'jump's. For each jump, the three previously chosen tiles were examined to see if they were neighbors. If so, the sequence was included in the analysis; otherwise the sequence was left out. Such a 'lookback' of 3 choices before a jump to characterize sequences was selected because more than 3 resulted in very few choice sequences, less power, and many fewer subjects, whereas fewer than 3 included many incidental two-choice sequences. Next, every information or reward outcome from every choice (whether part of the sequence or not) was

computed to find the average across all choices. Reward outcomes were determined simply on the basis of whether a reward was received or not. Information outcomes were determined by taking the difference between Shannon entropies of the distribution before and after a choice outcome. Finally, the average information (Fig. 3A) or reward (Fig. 3B) for each choice in the sequences as well as the global average across all choices was plotted, revealing the evidence for an average intake threshold rule.

Subjects' ability to forage was assessed using a custom 'forager score'. Foraging refers to decisions made in a sequential, non-exclusive, accept-or-reject context where options occur one at a time, foragers can accept or reject them, and rejected options can be returned to [120, 121]. A classic foraging decision is patch leaving where foragers must decide whether to continue foraging at a resource patch or to leave a patch to search for a new one. In the battleship task, we operationally defined a resource as a connected set of tiles, and the decision to leave a resource was defined as a jump. We consider three features of choice sequences that reflect foraging (inspired by [47, 70, 105]). First, while subjects decide to stay at a resource, the information or reward gained from choice outcomes should be above the average for the environment. Second, choice outcomes prior to deciding to leave a resource should be below average. Finally third (and as a consequence of the first two), outcomes preceding stay decisions should be above those preceding leave decisions. We constructed a forager score on the basis of these three features. For the kth information outcome k_i, pre-jump information outcome j_i, average information outcome $\bar{I}$, and subject s, let the forager score F_s be

$$F_s = (\sum (k_i > \bar{I}) + (j_i < \bar{I}) + \sum (k_i > j_i)) / 5,$$

where x > y is 1 if true and 0 if false. A score of 5 perfectly matches a foraging pattern of choices (two choice outcomes prior to pre-jump above average + pre-jump outcome below average + two choice outcomes prior to pre-jump above pre-jump outcome). This score was then regressed against the end of learning as determined by the final changepoint (Fig. 4). As a control for this analysis, a similar score was constructed for rewards that used the reward outcomes following each choice and the average reward outcomes across all choices.

References

1. Maia, T.V., *Reinforcement learning, conditioning, and the brain: Successes and challenges.* Cognitive, Affective, & Behavioral Neuroscience, 2009. **9**(4): p. 343-364.
2. Braun, D.A., et al., *Structure learning in action.* Behavioural brain research, 2010. **206**(2): p. 157-165.
3. Gershman, S.J. and Y. Niv, *Learning latent structure: carving nature at its joints.* Current opinion in neurobiology, 2010. **20**(2): p. 251-256.
4. Wilson, R.C. and Y. Niv, *Inferring relevance in a changing world.* Frontiers in human neuroscience, 2012. **5**: p. 189.
5. Dayan, P. and K.C. Berridge, *Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation.* Cognitive, Affective, & Behavioral Neuroscience, 2014. **14**(2): p. 473-492.

6. Gershman, S.J., et al., *Discovering latent causes in reinforcement learning.* Current Opinion in Behavioral Sciences, 2015. **5**: p. 43-50.
7. Tervo, D.G.R., et al., *Toward the neural implementation of structure learning.* Current opinion in neurobiology, 2016. **37**: p. 99-105.
8. Sutton, R.S. and A.G. Barto, *Reinforcement learning : an introduction*. Adaptive computation and machine learning. 1998, Cambridge, Mass.: MIT Press. xviii, 322 p.
9. Lee, D., et al., *Neural Basis of Reinforcement Learning and Decision Making.* Annual Review of Neuroscience, 2012. **35**(1): p. 287-308.
10. Daw, N.D., et al., *Model-based influences on humans' choices and striatal prediction errors.* Neuron, 2011. **69**(6): p. 1204-15.
11. Settles, B., *Active learning.* Synthesis Lectures on Artificial Intelligence and Machine Learning, 2012. **6**(1): p. 1-114.
12. Holland, J.H., *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. 1992: MIT press.
13. Schmidhuber, J. *Curious model-building control systems*. in *1991 IEEE International Joint Conference on Neural Networks.* 1991. IEEE.
14. Thrun, S. and K. Möller. *Active exploration in dynamic environments*. in *Advances in neural information processing systems*. 1992.
15. Thrun, S., *Exploration in active learning.* Handbook of Brain Science and Neural Networks, 1995: p. 381-384.
16. Cohen, J.D., et al., *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 B: Biological Sciences, 2007. **362**(1481): p. 933-942.
17. Kemp, C. and J.B. Tenenbaum, *Structured statistical models of inductive reasoning.* Psychological review, 2009. **116**(1): p. 20.
18. Griffiths, T.L. and J.B. Tenenbaum, *Structure and strength in causal induction.* Cognitive psychology, 2005. **51**(4): p. 334-384.
19. Markant, D.B. and T.M. Gureckis, *Is it better to select or to receive? Learning via active and passive hypothesis testing.* Journal of Experimental Psychology: General, 2014. **143**(1): p. 94.
20. Gureckis, T. and D. Markant. *Active learning strategies in a spatial concept learning game*. in *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2009.
21. Markant, D. and T. Gureckis. *Does the utility of information influence sampling behavior?* in *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2012.
22. Redish, A.D., et al., *Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling.* Psychol Rev, 2007. **114**(3): p. 784-805.
23. Tenenbaum, J.B., et al., *How to grow a mind: Statistics, structure, and abstraction.* science, 2011. **331**(6022): p. 1279-1285.
24. Collins, A.G. and M.J. Frank, *Cognitive control over learning: Creating, clustering, and generalizing task-set structure.* Psychological review, 2013. **120**(1): p. 190.
25. Collins, A.G., et al., *Human EEG uncovers latent generalizable rule structure during learning.* Journal of Neuroscience, 2014. **34**(13): p. 4677-4685.

26. Koechlin, E., *An evolutionary computational theory of prefrontal executive function in decision-making.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1655).
27. Niv, Y., et al., *Reinforcement learning in multidimensional environments relies on attention mechanisms.* Journal of Neuroscience, 2015. **35**(21): p. 8145-8157.
28. Schuck, N.W., et al., *Human orbitofrontal cortex represents a cognitive map of state space.* Neuron, 2016. **91**(6): p. 1402-1412.
29. Tomov, M.S., et al., *Neural Computations Underlying Causal Structure Learning.* Journal of Neuroscience, 2018: p. 3336-17.
30. Saez, A., et al., *Abstract context representations in primate amygdala and prefrontal cortex.* Neuron, 2015. **87**(4): p. 869-881.
31. Cheng, P.W., *From Covariation to causation: a causal power theory.* Psychol Rev, 1997. **104**(2): p. 367-405.
32. Glymour, C., *Learning causes: Psychological explanations of causal explanation.* Minds and machines, 1998. **8**(1): p. 39-60.
33. Pearl, J., *Causality*. 2009: Cambridge university press.
34. Tenenbaum, J.B. and T.L. Griffiths, *Generalization, similarity, and Bayesian inference.* Behavioral and brain sciences, 2001. **24**(4): p. 629-640.
35. Glymour, C., *Learning, prediction and causal Bayes nets.* Trends in cognitive sciences, 2003. **7**(1): p. 43-48.
36. Steyvers, M., et al., *Inferring causal networks from observations and interventions.* Cognitive science, 2003. **27**(3): p. 453-489.
37. Courville, A.C., et al. *Similarity and discrimination in classical conditioning: A latent variable account*. in *Advances in neural information processing systems*. 2005.
38. Gopnik, A., et al., *A theory of causal learning in children: causal maps and Bayes nets.* Psychological review, 2004. **111**(1): p. 3.
39. Sloman, S., *Causal models: How people think about the world and its alternatives*. 2005: Oxford University Press.
40. Tenenbaum, J.B., et al., *Theory-based Bayesian models of inductive learning and reasoning.* Trends in cognitive sciences, 2006. **10**(7): p. 309-318.
41. Courville, A.C., et al., *Bayesian theories of conditioning in a changing world.* Trends Cogn Sci, 2006. **10**(7): p. 294-300.
42. Gershman, S.J., *Context-dependent learning and causal structure.* Psychonomic bulletin & review, 2017. **24**(2): p. 557-565.
43. Waldmann, M., *The Oxford handbook of causal reasoning*. 2017: Oxford University Press.
44. Wilson, R.C., et al., *Orbitofrontal cortex as a cognitive map of task space.* Neuron, 2014. **81**(2): p. 267-79.
45. Niv, Y., *Learning task-state representations.* Nature neuroscience, 2019. **22**(10): p. 1544-1553.
46. Gershman, S.J., et al., *Context, learning, and extinction.* Psychol Rev, 2010. **117**(1): p. 197-209.
47. Stephens, D.W. and J.R. Krebs, *Foraging Theory*. 1986, Princeton, NJ: Princeton University Press.
48. Osu, R., et al., *Practice reduces task relevant variance modulation and forms nominal trajectory.* Scientific reports, 2015. **5**(1): p. 1-17.

49. Gallistel, C.R., et al., *The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect.* Journal of Experimental Psychology: Animal Behavior Processes, 2001. **27**(4): p. 354.
50. Inclan, C. and G.C. Tiao, *Use of cumulative sums of squares for retrospective detection of changes of variance.* Journal of the American Statistical Association, 1994. **89**(427): p. 913-923.
51. Rescorla, R.A. and A.R. Wagner, *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement*, in *Classical Conditioning II: Current Research and Theory*, A.H. Black and W.F. Prokasy, Editors. 1972, Appleton-Century-Crofts: New York.
52. Minsky, M., *Steps toward artificial intelligence.* Proceedings of the IRE, 1961. **49**(1): p. 8-30.
53. Kaelbling, L.P., et al., *Reinforcement Learning: A Survey.* Journal of Artificial Intelligence Research, 1996. **4**: p. 237-285.
54. Sutton, R.S., et al., *Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning.* Artificial intelligence, 1999. **112**(1-2): p. 181-211.
55. Koechlin, E., *Prefrontal executive function and adaptive behavior in complex environments.* Current opinion in neurobiology, 2016. **37**: p. 1-6.
56. Addicott, M., et al., *A primer on foraging and the explore/exploit trade-off for psychiatry research.* Neuropsychopharmacology, 2017. **42**(10).
57. Shannon, C.E. and W. Weaver, *The Mathematical Theory of Communication*. 1963, Urbana and Chicago: University of Illinois Press.
58. Crupi, V., et al., *Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search.* Cognitive Science, 2018. **42**(5): p. 1410-1456.
59. Gregory, R., *On how little information controls so much behaviour.* Ergonomics, 1970. **13**(1): p. 25-35.
60. Snyder, M. and W.B. Swann, *Hypothesis-testing processes in social interaction.* Journal of Personality and Social Psychology, 1978. **36**(11): p. 1202.
61. Trope, Y. and M. Bassok, *Confirmatory and diagnosing strategies in social information gathering.* Journal of personality and social psychology, 1982. **43**(1): p. 22.
62. Trope, Y. and A. Liberman, *Social hypothesis testing: Cognitive and motivational mechanisms.* 1996.
63. Wason, P.C., *Reasoning*, in *New Horizons in Psychology*, B. Foss, Editor. 1966. p. 135-151.
64. Wason, P.C., *Reasoning about a rule.* Quarterly journal of experimental psychology, 1968. **20**(3): p. 273-281.
65. Klayman, J. and Y.-W. Ha, *Confirmation, disconfirmation, and information in hypothesis testing.* Psychological review, 1987. **94**(2): p. 211.
66. Siskind, J.M., *A computational study of cross-situational techniques for learning word-to-meaning mappings.* Cognition, 1996. **61**(1-2): p. 39-91.
67. Poletiek, F.H., *Hypothesis-testing behaviour*. 2013: Psychology Press.
68. Markant, D.B., et al., *Self-directed learning favors local, rather than global, uncertainty.* Cognitive science, 2016. **40**(1): p. 100-120.
69. Pirolli, P. and S. Card, *Information foraging.* Psychological review, 1999. **106**(4): p. 643.

70. Fu, W.-T. and P. Pirolli, *SNIF-ACT: A cognitive model of user navigation on the World Wide Web.* Human–Computer Interaction, 2007. **22**(4): p. 355-412.
71. Pirolli, P.L.T., *Information foraging theory: Adaptive interaction with information*. 2007: Oxford University Press.
72. Najemnik, J. and W.S. Geisler, *Optimal eye movement strategies in visual search.* Nature, 2005. **434**(7031): p. 387-391.
73. Vergassola, M., et al., *'Infotaxis' as a strategy for searching without gradients.* Nature, 2007. **445**(7126): p. 406.
74. Johnson, A., et al., *The hippocampus and exploration: dynamically evolving behavior and neural representations.* Frontiers in human neuroscience, 2012. **6**.
75. Manohar, S.G. and M. Husain, *Attention as foraging for information and value.* Frontiers in human neuroscience, 2013. **7**: p. 711.
76. Oaksford, M. and N. Chater, *A rational analysis of the selection task as optimal data selection.* Psychological Review, 1994. **101**(4): p. 608.
77. Markant, D. and T. Gureckis. *Modeling information sampling over the course of learning*. in *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2011.
78. Tsividis, P., et al. *Information selection in noisy environments with large action spaces*. in *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2014.
79. Markant, D. and T. Gureckis. *Category learning through active sampling*. in *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2010.
80. Rich, A.S. and T.M. Gureckis, *Exploratory Choice Reflects the Future Value of Information.* Decision, 2017.
81. Miller, G., *Informavores*, in *The study of information: Interdisciplinary messages*, F. Machlup and U. Mansfield, Editors. 1983, Wiley-Interscience. p. 111-113.
82. Oaksford, M. and N. Chater, *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. 1998: Psychology Press/Erlbaum (UK) Taylor & Francis.
83. Oaksford, M. and N. Chater, *Bayesian rationality: The probabilistic approach to human reasoning*. 2007: Oxford University Press.
84. Nelson, J.D., *Finding useful questions: on Bayesian diagnosticity, probability, impact, and information gain.* Psychological review, 2005. **112**(4): p. 979.
85. Nelson, J.D., et al., *Experience matters: Information acquisition optimizes probability gain.* Psychological science, 2010. **21**(7): p. 960-969.
86. Nelson, J.D., et al., *Children's sequential information search is sensitive to environmental probabilities.* Cognition, 2014. **130**(1): p. 74-80.
87. Johnson-Laird, P.N. and P.C. Wason, *A theoretical analysis of insight into a reasoning task.* Cognitive psychology, 1970. **1**(2): p. 134-148.
88. Coenen, A., et al., *Asking the right questions about the psychology of human inquiry: Nine open challenges.* Psychonomic Bulletin & Review, 2019. **26**(5): p. 1548-1587.
89. Nelson, J.D., et al., *Towards a theory of heuristic and optimal planning for sequential information search.* 2018.
90. Meder, B., et al., *Stepwise versus globally optimal search in children and adults.* Cognition, 2019. **191**: p. 103965.

91. Bramley, N.R., et al., *Conservative forgetful scholars: How people learn causal structure through sequences of interventions.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 2015. **41**(3): p. 708.
92. Meier, K.M. and M.R. Blair, *Waiting and weighting: Information sampling is a balance between efficiency and error-reduction.* Cognition, 2013. **126**(2): p. 319-325.
93. Wyckoff Jr, L.B., *The role of observing responses in discrimination learning. Part I.* Psychological review, 1952. **59**(6): p. 431.
94. Roper, K.L. and T.R. Zentall, *Observing behavior in pigeons: The effect of reinforcement probability and response cost using a symmetrical choice procedure.* Learning and Motivation, 1999. **30**(3): p. 201-220.
95. Bromberg-Martin, E.S. and O. Hikosaka, *Midbrain dopamine neurons signal preference for advance information about upcoming rewards.* Neuron, 2009. **63**(1): p. 119-26.
96. Blanchard, T.C., et al., *Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity.* Neuron, 2015. **85**(3): p. 602-614.
97. Kreps, D.M. and E.L. Porteus, *Temporal resolution of uncertainty and dynamic choice theory.* Econometrica: journal of the Econometric Society, 1978: p. 185-200.
98. Beierholm, U.R. and P. Dayan, *Pavlovian-instrumental interaction in 'observing behavior'.* PLoS computational biology, 2010. **6**(9): p. e1000903.
99. Foley, N.C., et al., *Parietal neurons encode expected gains in instrumental information.* Proceedings of the National Academy of Sciences, 2017. **114**(16): p. E3315-E3323.
100. Horan, M., et al., *Parietal neurons encode information sampling based on decision uncertainty.* Nature neuroscience, 2019. **22**(8): p. 1327-1335.
101. Viswanathan, G.M., et al., *The physics of foraging: an introduction to random searches and biological encounters*. 2011: Cambridge University Press.
102. Hills, T.T., *Animal Foraging and the Evolution of Goal-Directed Cognition.* Cognitive Science, 2006. **30**(1): p. 3-41.
103. Hills, T.T., et al., *Adaptive Lévy Processes and Area-Restricted Search in Human Foraging.* PLOS ONE, 2013. **8**(4): p. e60488.
104. Todd, P.M. and T.T. Hills, *Foraging in mind.* Current Directions in Psychological Science, 2020. **29**(3): p. 309-315.
105. Charnov, E.L., *Optimal foraging, the marginal value theorem.* Theor Popul Biol, 1976. **9**(2): p. 129-36.
106. Hills, T.T., et al., *The central executive as a search process: priming exploration and exploitation across domains.* Journal of Experimental Psychology: General, 2010. **139**(4): p. 590.
107. Huberman, B.A., et al., *Strong regularities in world wide web surfing.* Science, 1998. **280**(5360): p. 95-97.
108. Pirolli, P., *Rational analyses of information foraging on the web.* Cognitive science, 2005. **29**(3): p. 343-373.
109. Cain, M.S., et al., *A Bayesian optimal foraging model of human visual search.* Psychological Science, 2012: p. 0956797612440460.
110. Wolfe, J.M., *When is it time to move to the next raspberry bush? Foraging rules in human visual search.* Journal of vision, 2013. **13**(3): p. 1-17.

111. Payne, S.J., et al., *Discretionary task interleaving: heuristics for time allocation in cognitive foraging.* Journal of Experimental Psychology: General, 2007. **136**(3): p. 370.
112. Wilke, A., et al., *Fishing for the right words: Decision rules for human foraging behavior in internal search tasks.* Cognitive Science, 2009. **33**(3): p. 497-529.
113. Metcalfe, J. and W.J. Jacobs, *People's study time allocation and its relation to animal foraging.* Behavioural processes, 2010. **83**(2): p. 213-221.
114. Payne, S. and G. Duggan, *Giving up problem solving.* Memory & Cognition, 2011. **39**(5): p. 902-913.
115. Hills, T.T., et al., *Optimal foraging in semantic memory.* Psychological review, 2012. **119**(2): p. 431.
116. Hills, T.T., et al., *Foraging in semantic fields: How we search through memory.* Topics in cognitive science, 2015. **7**(3): p. 513-534.
117. Turrin, C., et al., *Social resource foraging is guided by the principles of the Marginal Value Theorem.* Scientific Reports, 2017. **7**(1): p. 11274.
118. Rothe, A., et al. *Asking and evaluating natural language questions*. in *CogSci*. 2016.
119. Gottlieb, J. and P.-Y. Oudeyer, *Towards a neuroscience of active sampling and curiosity.* Nature Reviews Neuroscience, 2018. **19**(12): p. 758-770.
120. Calhoun, A.J. and B.Y. Hayden, *The foraging brain.* Current Opinion in Behavioral Sciences, 2015. **5**: p. 24-31.
121. Barack, D.L. and M.L. Platt, *Engaging and Exploring: Cortical Circuits for Adaptive Foraging Decisions*, in *Impulsivity*. 2017, Springer. p. 163-199.