Visuospatial information foraging describes search behavior in learning latent environmental features

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

* Corresponding author: dbarack@gmail.com

Abstract

In the real world, making sequences of decisions to achieve goals often depends upon the ability to learn aspects of the environment that are not directly perceptible. Learning these so-called latent features requires seeking information about them, a process distinct from learning about reinforcement contingencies. Prior efforts to study latent feature learning often use single decisions, use few features, and fail to distinguish between reinforcement-seeking and information-seeking. To overcome this, we designed a task in which humans and monkeys made a series of choices to search for shapes hidden on a grid. Reward and information outcomes from uncovering parts of shapes were not perfectly correlated and their effects could be disentangled. Members of both species adeptly learned the shapes. Both species preferred to select informative tiles earlier in trials than rewarding ones, searching a part of the grid until their outcomes dropped below the average information outcome–a pattern consistent with foraging behavior. In addition, how quickly humans learned the shapes was predicted by how well their choice sequences matched the foraging pattern. This adaptive search for information may underlie the ability in humans and monkeys to learn latent features to support goal-directed behavior in the long run.

Introduction

All animals must learn the features of their environments and how they tend to occur together to make better choices to achieve their goals. For example, suppose you are getting ready to leave for school and you need to find your backpack, which is somewhere around the house, perhaps in the closet or on the floor under your coat. While some features of the environment, like the color, texture and shape of the backpack, are perceptible, other latent features are not. Latent features are defined as those that must be inferred from perceptible features and that can manifest in a variety of perceptual ways. The rumpled coat on the floor can serve as a clue for a lost backpack, hidden underneath. The appearance of the coat, a perceptible feature, provides evidence for the underlying hidden bag, a latent feature. Such perceptual clues can help animals select actions that will more efficiently achieve their goals, such as getting going in the morning. Many contexts involve learning latent features on the basis of different patterns of visuospatial perceptions, ranging from the mundane, such as determining the ripeness of a fruit [1], to the esoteric, such as how to play a video game [2].

Despite the relevance to many aspects of cognition, how humans and other animals learn these latent features is only recently a focus of research in psychology and neuroscience. Many studies in psychology and related disciplines investigate how animals can learn features. While perceptible features in the environment can be learned by observation and directly reinforced by the outcomes of an individual's choices, latent features in environments must be inferred [3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. These latent features often capture the statistical structure of situations in the environment across stimuli, actions, time, space and a range of variables internal to the organism such as mood or cognitive state [13].

Latent features can be learned from the outcomes of choices. Outcome-motivated learning of latent features is typically studied by using rewards such as food or water to provide feedback that can be used to augment behavior [14, 15, 16]. In the real world, however, many behaviors go unrewarded and yet learning still occurs [17]; for example, there is no external reward associated with finding your backpack. Focusing only on rewards, then, fails to fully capture learning. Besides reward, the outcomes of choices also provide information, here understood as changes in the probabilities of different features or causes appearing in the environment. This information can also be used for learning; specifically, the acquisition of information can help construct or identify a latent feature. Since knowledge of latent features can be critical for developing and applying

efficient strategies to obtain reward and avoid aversive outcomes, a quest for information can itself be a motivating force for behavior. Understanding how reward-based and information-based motivations impact learning is consequently important for understanding how learning proceeds in the messy complexity of real-world environments.

Learning about a latent feature involves gathering reward or information from many decisions across different timescales. On a short timescale, such as individual choices, reward or information can reinforce actions and contingencies that have recently occurred. On a longer timescale, such as sequences of choices or even multiple episodes, linking together multiple outcomes can reinforce actions and contingencies that are statistically related. In general, recent attempts to study information gathering for latent feature learning have suffered from designs that either cannot fully distinguish between reward-driven and information-driven strategies despite requiring many decisions [18, 19, 20, 21], fail to investigate gathering information that can be used to learn latent features [22, 23, 24, 25, 26, 27], or fail to investigate the motivations behind sequences of decisions to gather information [28, 29].

To understand the role of reward and information outcomes in learning latent features over multiple decisions, we developed a novel behavioral paradigm based on the board game Battleship. On each trial, participants started with a grid of unchosen tiles, selected tiles from the grid to reveal whether a piece of a shape is hidden beneath the tile, and ended trials when all of a hidden shape's tiles had been revealed. Revealing a filled or empty square provided information about both the current trial's shape as well as about the set of possible shapes that could occur across trials. On the current trial, revealing a filled or empty square provides evidence for the hidden shape. Over many trials, participants learn which shapes out of thousands of possibilities could occur. In our task, information (changes in the probabilities of different shapes) is partly decorrelated from reward earned from turning over a tile (points for humans or squirts of juice for monkeys). Hence, the effect of information and reward outcomes on patterns of choices can be disambiguated in our task.

Humans are hypothesized to be exquisite information gatherers [30], but the degree to which this skill extends across the primate clade is unknown. To investigate information-motivated learning, we had both humans and monkeys perform our task. We observed that participants from both species were very adept at the Battleship-based task, quickly learning the shapes. Analysis of choice behavior revealed that tiles that were expected to be informative about the hidden shape

tended to be selected earlier in a trial, whereas tiles that had been rewarding in the past tended to be selected later. Importantly, over multiple choices, exploratory analysis revealed that both humans and monkeys tended to abandon local searches in one part of the grid to jump to a new part when their most recent outcome provided less information than the average for the environment, a type of foraging behavior previously observed for gathering rewards across species. Furthermore, the degree to which the patterns of choices of human participants matched a foraging pattern predicted how quickly shapes were learned. After learning, humans abandoned this pattern of information foraging in favor of gathering rewards. In contrast, monkeys continued to forage for information late in sessions. In sum, evidence from behavior on our task suggests both humans and monkeys searched for information across different timescales to learn shapes and used foraging computations to decide where to sample spatially on the grid to gather information.

Results

To study how animals learn latent features over multiple decisions, we designed a shape search task. The task contained thousands of possible latent features—the hidden shapes, sets of connected filled tiles at a location—and identification of these features could facilitate more rapid successful search. The task bears similarity to the board game 'Battleship'. Participants viewed a 5x5 tile grid and locations on the grid could be chosen to reveal either a filled or empty tile. On every trial, one of five shapes was pseudorandomly chosen. The shapes partly overlapped; for example, the 'H' shape overlapped with backwards-'L' shape at the bottom row, middle tile (Fig. 1A, bottom). Hence, uncovering a piece of the shape did not always reveal the identity of the shape for that trial.

Participants uncovered the shape over multiple choices by selecting tiles (Fig. 1A). At trial start, participants made a movement to a target at the center of the screen (humans: mouse-over; monkeys: saccade) until it disappeared after a variable delay. Participants then had unlimited time to choose a target. After a choice, if participants uncovered a filled tile ('hit'), then they received a reward (points for humans, juice for monkeys) before proceeding to the next choice. If participants failed to uncover a filled tile ('miss'), then they proceeded to the next choice with no reward. After each choice outcome and an inter-choice interval, the fixation point reappeared, and the sequence repeated until the shape was fully revealed (see methods). Trials ended once all filled

tiles that were part of the shape were uncovered. A trial, then, is the set of choices and outcomes from initial fixation with a fully occluded shape to the last choice that finished revealing the shape.

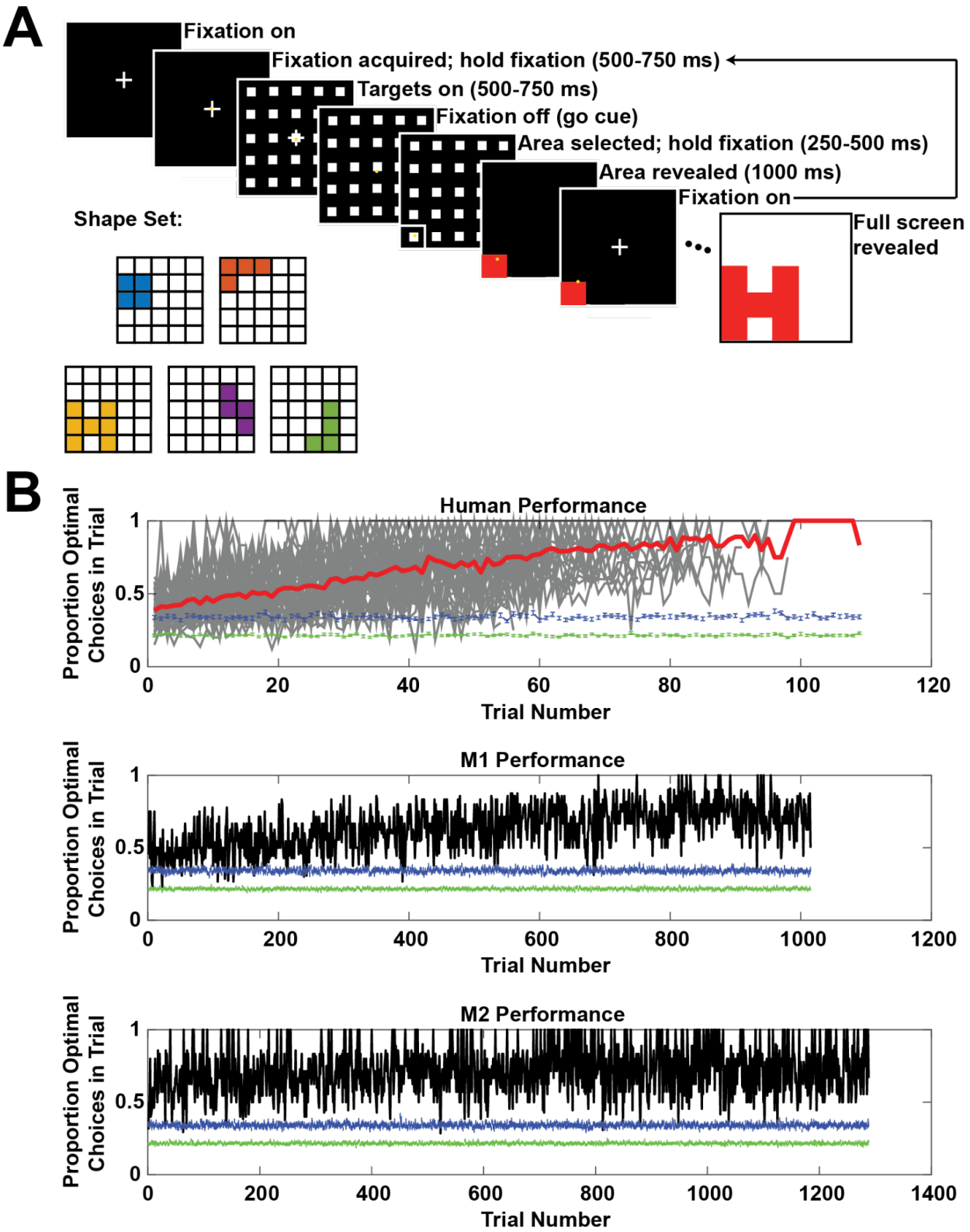

Figure 1. A. *Shape search task. The variable time intervals reported in the figure are for the monkeys; the human participants had hold times of 500 ms for fixation, 500 ms after targets on, and 250 ms to register selection of a target.* ***B.*** *Performance on the shape search task. Top panel: thick red line = average human performance (n = 42), thin gray lines = individual subject traces; middle panel: M1; bottom panel: M2. For all three panels, jagged green line at bottom is the average performance of a choice algorithm that randomly chooses tiles (100 iterations), and jagged blue line at bottom is the average performance of a choice algorithm that randomly chooses tiles until a hit and then performs a local area search (100 iterations). Points are mean ± s.e.m.*

To visualize performance, we plotted the proportion of choices that maximized rewards as a function of trial number (humans: 45 min session with as many trials performed as possible; M1 and M2: untimed sessions concatenated across 7 days). At each choice, some subset of the five shapes remain and, since the shapes overlapped, participants that choose tiles with the most overlap among the remaining shapes will maximize rewards. This strategy represents the Bellman optimal policy for selecting squares [14]. The proportion of choices that maximized reward increased over sessions for both species, with human participants (n = 42) increasing about an order of magnitude faster than monkeys in stabilizing the proportion of choices that maximized reward per trial around 70% (Fig. 1B). All participants outperformed both an algorithm that randomly selected tiles on every choice (Fig. 1B, green lines below data) and an algorithm that randomly selected tiles until a hit and searched locally thereafter (Fig. 1B, blue lines below data). The remainder of this paper will focus on analyses of this learning behavior.

Did rewards or information drive choices of individual tiles? Outcomes from each choice deliver both rewards and information about the hidden shape. In principle both reward and information could influence subsequent choices. To disentangle the relative contribution of reward and information, we computed the expected reward outcome and expected information outcome for each tile, updated after each choice made by participants. For each tile, the expected reward is defined by the number of hits from the choice of that tile divided by the number of times the tile was previously chosen. By contrast, expected information is computed by quantifying the expected change due to a hit or a miss in the entropy of the probability distribution characterizing the possible remaining shapes (see methods). Informally, one can think of these distributions as a set of probabilities, one for each shape. This distribution is updated during trials based on getting

hits or misses and updated after the end of the trial based on which shape was observed. Expected reward and expected information across choices were not significantly correlated (humans: $\rho = -0.15$; both monkeys: $\rho = -0.09$). A multinomial logistic regression was performed to regress choice number in trial against trial number in session, the expected information for the chosen tile, and expected reward for the chosen tile (Fig. 2). For human participants, the mean coefficient for expected information was positive and significantly different from the negative mean coefficient for expected reward ($t(df=14) = 6.18$, $p < 1x10^{-4}$; mean $\beta_{info} = 1.37 \pm 0.41$; mean $\beta_{reward} = -0.19 \pm 0.14$). This implies that in humans, information correlated with higher probability of earlier choices of tiles whereas reward correlated with higher probability of later choices of tiles. Paired t-tests of expected information and expected reward coefficients for each choice number revealed that information influenced choice significantly more than reward ($t(df=14) = 6.21$, $p < 5x10^{-5}$), 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. M2 was only marginally more driven by information ($t(df=7) = -2.03$, $p = 0.0822$) but significantly more driven by reward than M1 ($t(df=7) = -2.96$, $p < 0.05$). The influence of information on choice was greater than reward in M1 ($t(df=7) = 3.66$, $p < 0.01$) and M2 ($t(df=7) = 3.46$, $p < 0.05$).

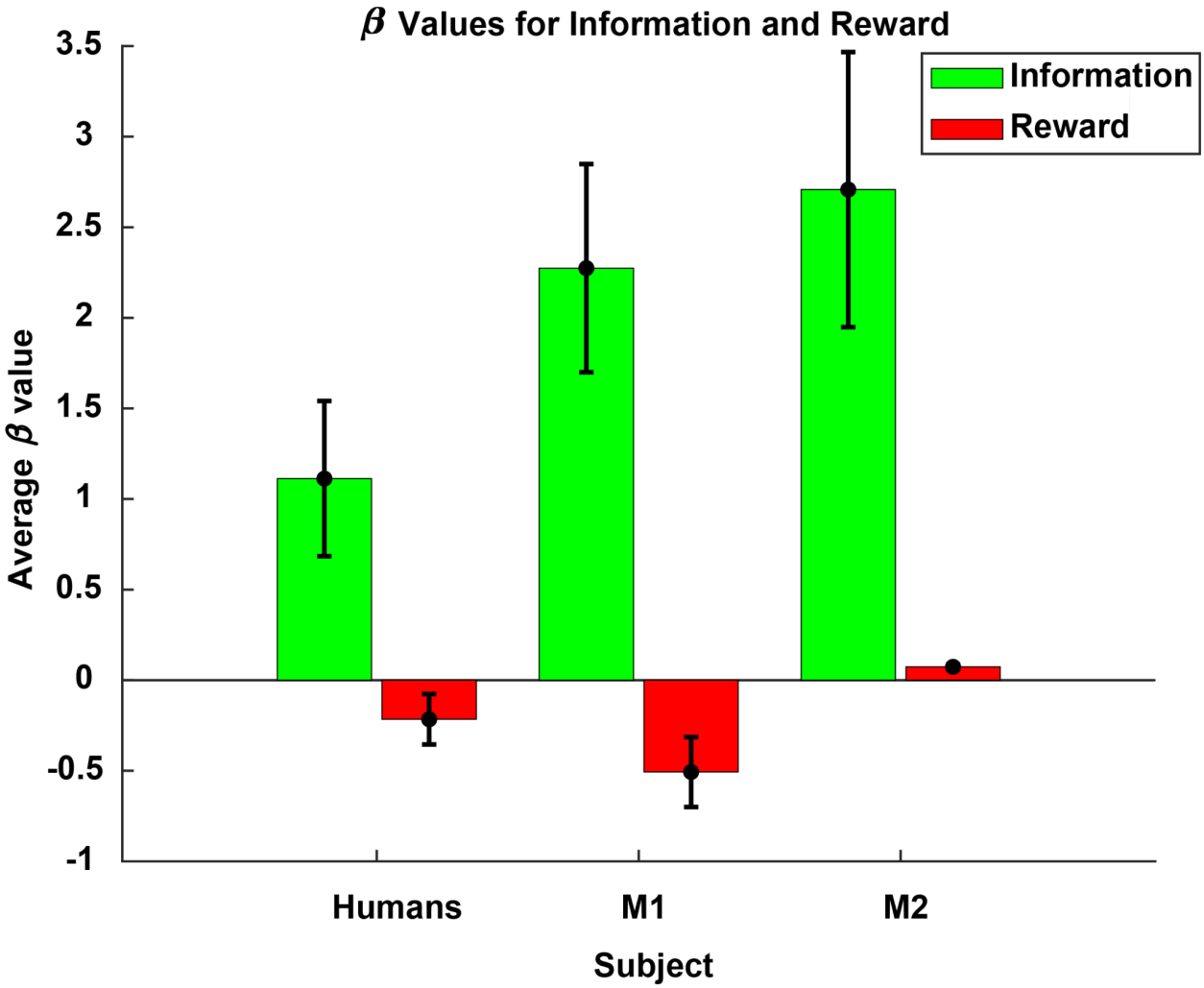

Figure 2. *The influence of expected information and expected reward on choice number in trial for humans, M1, and M2. Each column is the average beta (left pair of bars, across human participants and choice numbers, mean by subject ± 1 s.e.m.; middle and right pairs of bars, M1 and M2 respectively,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

After considering whether reward or information was driving participants' choices, we explored how participants made their choices, particularly in sequences. We wondered if participants explored the grid to learn the hidden shapes by choosing neighboring tiles. Humans and to a lesser extent monkeys increased their probability of choosing a neighboring tile after a hit over time (Fig. 3A, left column; OLS; humans, first row: $\beta = 0.0037 \pm 0.0002$ fraction of choices of neighboring tile after hit / trial, $p < 1\text{x}10^{-33}$; M1, second row: $\beta = 0.00013 \pm 0.000018$, $p < 1\text{x}10^{-10}$; M2, third row: $\beta = 0.000055 \pm 0.000013$, $p < 1\text{x}10^{-4}$). In contrast, the probability of choosing a

neighboring tile after a miss decreased over time for humans and for monkeys, albeit more weakly (Fig. 3A, right column; OLS; humans, first row: $\beta = -0.0014 \pm 0.0002$, $p < 1\text{x}10^{-14}$; M1, second row: $\beta = -0.000024 \pm 0.000013$, $p = 0.0695$; M2, third row: -0.000091 ± 0.0000093, $p < 1\text{x}10^{-20}$). In addition, following hits, humans showed a greater increase in their tendency to select neighboring tiles that were also hits than monkeys (Fig. 3C; OLS; Humans: $\beta = 0.0046 \pm 0.00024$ fraction chose neighboring hit after hit / trial, $p < 1\text{x}10^{-26}$; M1: $\beta = 0.00018 \pm 0.000020$ fraction of choices neighboring hit after hit / trial, $p < 1\text{x}10^{-16}$; M2: $\beta = 0.00014 \pm 0.000015$ fraction chose neighboring hit after hit / trial, $p < 1\text{x}10^{-9}$). In sum, while members of both species tended to more frequently choose hits following hits over time, humans increased their probability of doing so more quickly than monkeys.

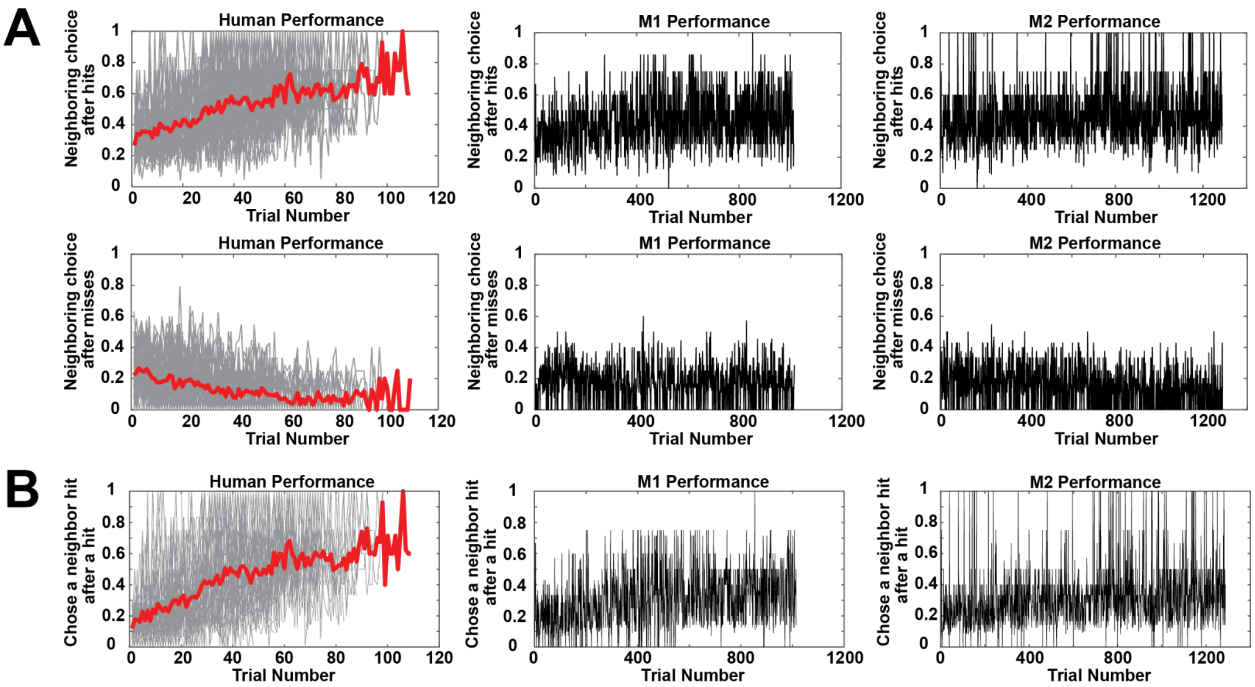

Figure 3. A. *Probability of choosing a neighboring tile after a hit (left column) or a miss (right column). First row: humans; second: M1; third: M2.* ***B.*** *Probability of choosing a neighboring hit after a hit. Left: humans; middle: M1; right: M2. All plots: thick red lines are average probability of choice; gray lines are individual human participant performance.*

What variables influenced these choices? We ran a mixed-effects binomial regression on human participants' choices with dependent variable of choice of neighboring tile (= 1) or not (= 0) and independent fixed-effect variables trial number in session, choice number in trial, information outcome from the previous choice, expected information for the current choice, reward outcome from the previous choice, and expected reward for the current choice and independent random-

effect variable of subject identity. All main effects were significant ($p < 0.05$, Bonferroni corrected). The effect of information and reward outcome was positive ($\beta_{info} = 1.75 \pm 0.069$; $\beta_{reward} = 1.34 \pm 0.056$), indicating that larger information or reward outcomes predicted choice of a neighboring tile. Larger information or reward outcomes tend to result from hits; since shapes were connected filled tiles, after a hit, some number of the neighboring tiles will typically also be parts of shapes, and so choosing a neighboring tile is rational. The effect of expected information and expected reward were negative ($\beta_{exp_info} = -0.87 \pm 0.091$; $\beta_{exp_reward} = -0.51 \pm 0.061$), indicating that larger expected information or reward predicted choice of a non-neighboring tile. Larger expected information or reward tend to result from misses; if a choice is a miss, then the probability that a shape is nearby is low, and so the next choice should be in a different part of the grid. The same regression was run on monkeys, which revealed significant main effects of trial number, expected information, and information and reward outcomes but not choice number or expected reward ($p < 0.05$, Bonferroni corrected). The sign of the significant main effects matched the human participants ($\beta_{info} = 0.86 \pm 0.0403$; $\beta_{reward} = 1.41 \pm 0.048$; $\beta_{exp_info} = -0.33 \pm 0.045$), consistent with similar motivations for choosing a neighboring tile or not. Participants were driven to search neighboring tiles by good outcomes and to search further away by bad ones.

We were primarily interested in how subjects explored the grid to learn the shapes, which we operationalized as the increase in their performance (Fig. 1B). To understand this, we need to define the learning period. Learning starts with the first trial. We reasoned that two main effects should be evident to mark the end of learning. First, the mean number of choices to finish revealing a shape should diminish during learning. Second, the variance in the number of choices should also diminish as participants learned the shapes (cf. [31]). To detect changes in these values, a changepoint detection test was run on the mean and variance of participants' choices across all shapes and trials (see methods) [32, 33]. The end of learning was set to the last changepoint (whether due to changes in mean or variance) across all shapes.

Looking at participants' sequence of choices when they did not choose neighboring tiles can reveal patterns of outcomes that aid in learning. Participants may search in a restricted area [34], hunting for shapes in a part of the grid. We operationalized such an area restricted search as the choice of at least three neighboring tiles in a row. We considered choices when participants made a decision to 'jump', a choice of a tile more than one tile away from the previous choice (i.e., the choice of a non-neighboring tile), after such area restricted search. A significant number of trials

contain these sequences (humans: 0.28 ± 0.012 fraction of trials; M1: 0.27; M2: 0.29). These jumps are the result of decisions to sample a new area of the grid.

Both human and monkey participants tended to jump after these sequences of choices (three neighboring tiles) when the information intake dropped below the average information intake during learning (Fig. 4A). This drop is the result of learning less about the current hidden shape, for example by getting misses that rule out fewer shapes. In humans, this pattern was not observed when these choices were plotted using reward outcomes (Fig. 4B, left panel). Monkeys' jumps were equally well-predicted by recent outcomes falling below the average information (Fig. 4A, middle and right panels) or below the average reward (Fig. 4B, middle and right panels). To explore the outcomes that drove this pattern of choice in humans, we separated the eight distinct sequences of hits and misses for three outcomes in a row by the last outcome in the sequence, either a hit or a miss, and plotted the information outcomes before jumping based only on last choice misses and last choice hits (Fig. 4C, left panel). Information foraging was driven primarily by misses: the pattern disappears if misses are left out but not if hits are left out. This pattern was observed in M1 but not M2 (Fig. 4C, middle and right panels respectively).

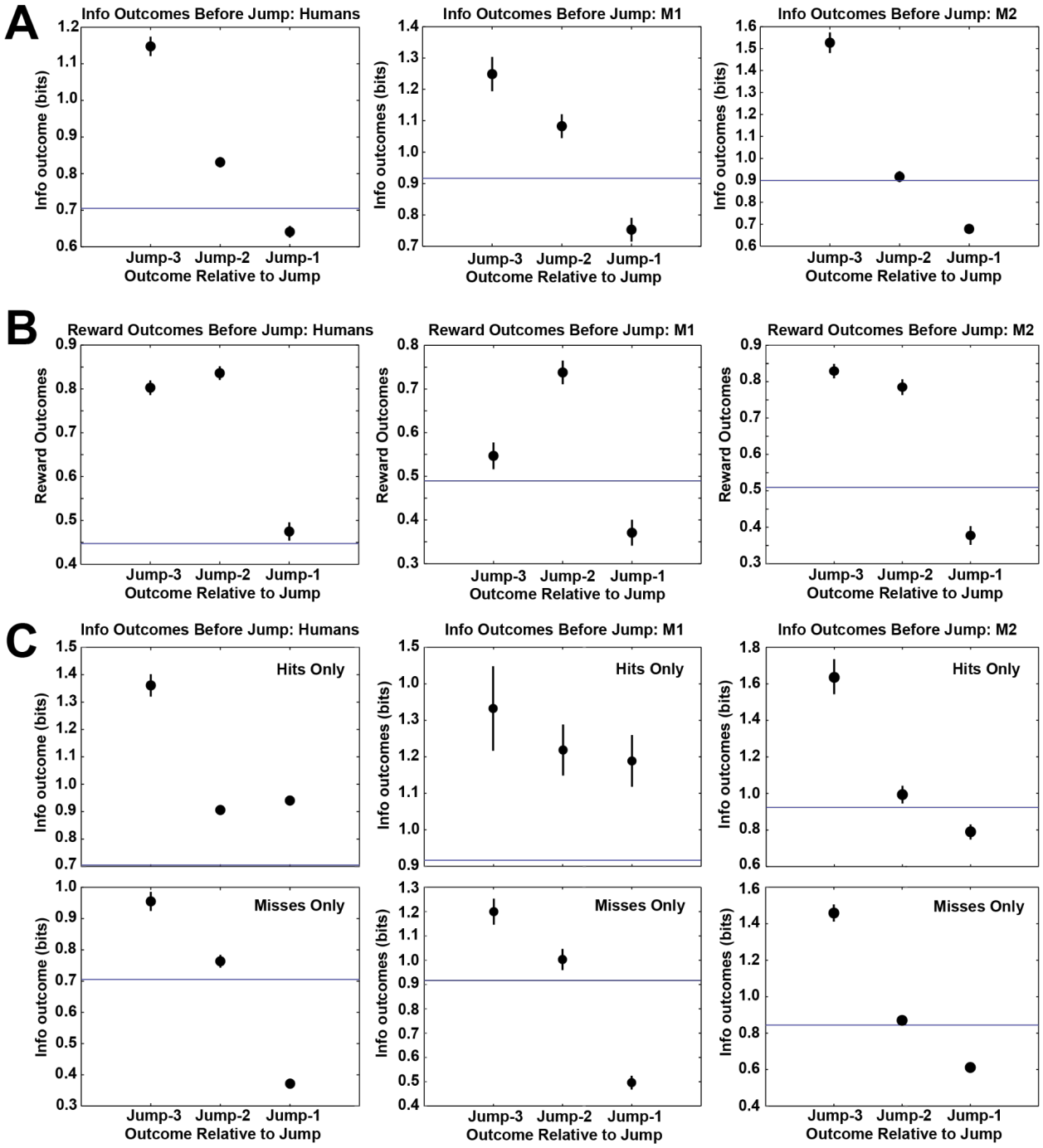

Figure 4. *Information and reward outcomes three choices before a jump, two choices before a jump, and the choice before a jump.* ***A.*** *Left: human information foraging; Middle: M1; Right: M2.* ***B.*** *Left: human reward foraging; Middle: M1; Right: M2. Note that the reward outcome for a choice before a jump for humans was not below the average reward outcome across all choices. Both* ***A*** *and* ***B*** *include all trials during learning only.* ***C.*** *Information outcomes three choices before a jump broken out by choice outcome just prior to jump, either 'Miss' (bottom panels) or 'Hit' (top panels).*

Humans: left pair of panels; M1: middle pair of panels; M2: right pair of panels. Points: mean, error bars: ± 1 s.e.m. Error bars sometimes occluded by points.

The unexpected observed pattern of choices, to jump to a different part of the grid when recent outcomes fall below the average, qualitatively matches foraging patterns of choices [35]. Animals can forage for a range of resources, including rewards [35], information [36], social interactions [37] and more. When foraging, animals seek to maximize the rate of intake of these resources. A simple rule for maximizing intake rates is to compare the current resource intake rate to either the average [38] or expected [39, 40] intake. When the current rate drops below that value, the forager should switch from exploiting the current location to exploring for a new one. On our task, participants' behavior matched this rule.

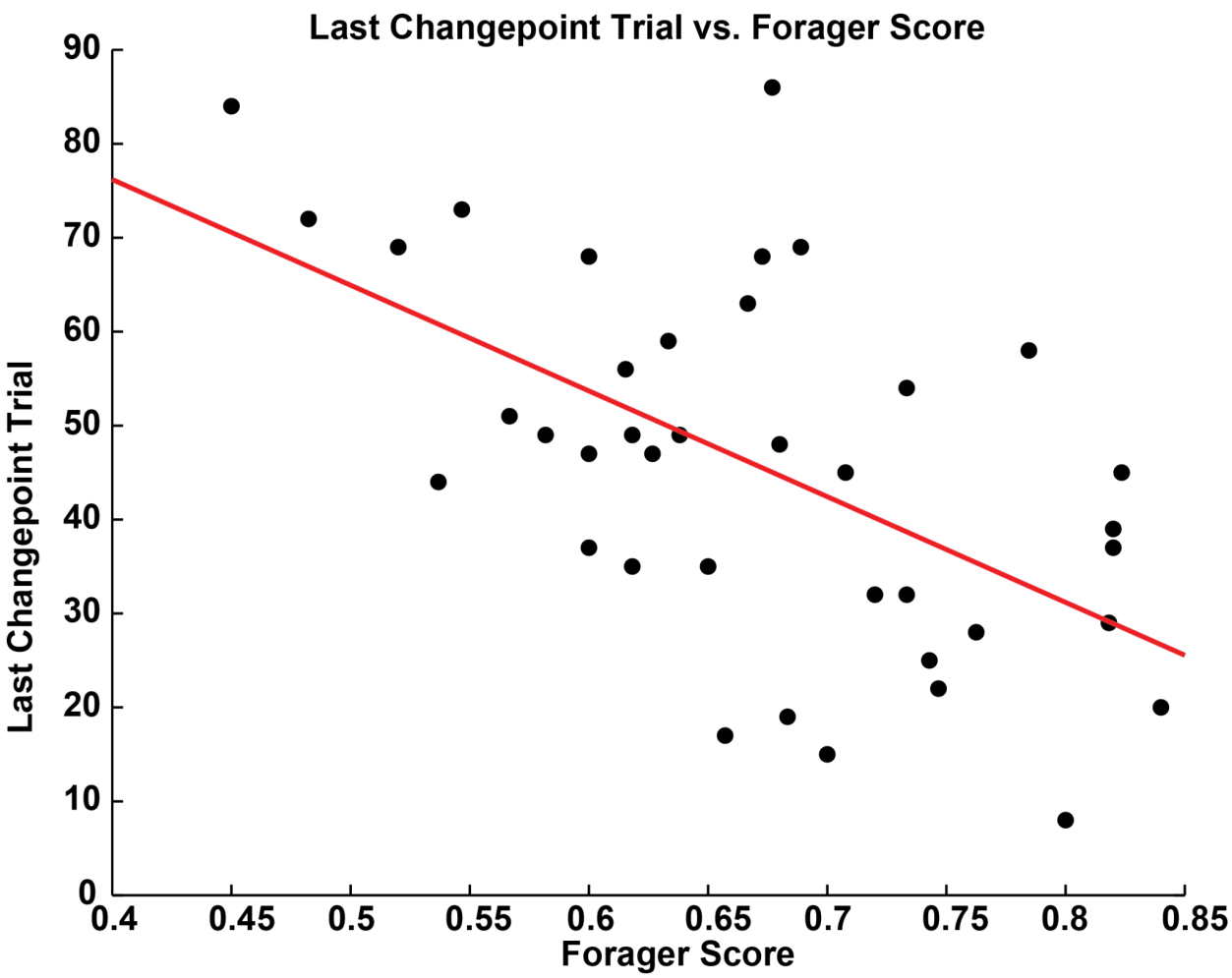

Figure 5. *For human participants, the information forager score F_s predicts the speed of learning. Each point is a single subject. Red line: OLS regression (β_{slope} = -112.54 ± 26.78 trials until changepoint/a.u. forager score, $p < 0.0005$).*

But why should humans or monkeys forage on our task? Does it help to learn the shapes? Answering this question required quantifying forager performance and comparing that performance to a measure such as how quickly the shapes were learned. To quantify forager performance, a forager score was computed for each subject for two periods: during and after learning (see methods). Better information foraging scores for humans during learning predicted earlier final changepoints and hence faster learning (Fig. 5; ordinary least squares (OLS), β_{slope} = -112.54 ± 26.78 trials until changepoint/a.u. forager score, $p < 0.0005$, $\rho = -0.57$). No such relationship was found for reward foraging scores (OLS, β_{slope} = 41.85 ± 43.99, $p > 0.3$, $\rho = 0.1545$, not plotted). During learning, human information forager scores (mean forager score F_S= 0.67 ± 0.016) were not significantly different from M1 (one-sample t-test; M1 = 0.67; t(df=38) = 0.36, $p > 0.7$) and marginally different from M2 (one-sample t-tests; M2 = 0.70; t(df=38) = -1.90, $p > 0.05$). After learning, humans (mean F_S = 0.56 ± 0.042) had significantly lower information foraging scores than both monkeys (one-sample t-tests; M1 = 0.72, t(df=37) = -3.82, $p < 5x10^{-4}$; M2 = 0.72, t(df=37) = -3.84, $p < 5x10^{-4}$). Further, humans foraged for information significantly more during learning compared to after shapes had been learned (paired t-test, t(df=75) = 2.63, $p < 0.05$) whereas monkeys marginally increased (M1: 0.67 during learning, 0.72 after; M2: 0.70 during, 0.72 after). In sum, better information foraging in humans predicted faster learning (as judged by their last changepoint) and humans ceased information foraging after learning.

Discussion

Learning environments with many latent features is a difficult challenge for any organism. To investigate how humans and monkeys learn these environments, we implemented a search task with many hidden shapes. We discovered that 1) for single choices, informative tiles were preferred earlier in trials than rewarding ones; 2) for pairs of choices, information outcomes predicted decisions to choose a neighboring tile; 3) for sequences of choices, decreases in information outcomes below the average of the environment predicted decisions to sample new areas of the grid, a signature of information foraging; and 4) the degree to which humans' choice sequences matched foraging patterns predicted learning.

Adaptive decision making in the real world requires learning latent features of the environment. Learning such features requires participants to make multiple decisions over extended periods of time in constantly changing environments and to keep track of outcomes across different

timescales. At shorter timescales, outcomes from individual choices provide momentary evidence about the current environment, such as revealing a hit or a miss. Series of outcomes from sequences of choices are at longer timescales and can fully reveal environmental features, such as a shape. At even longer timescales, multiple latent environmental features might be learned, such as many different shapes. Despite the importance of this critical cognitive skill, scant studies have investigated this type of learning at multiple timescales. Here, we probed for the first time in both humans and monkeys how latent features are learned in a temporally extended serial decision-making task where the environment constantly changes and outcomes must be tracked both during trials and across trials.

The drive to learn environments over many decisions can be motivated by the search for reward or for information. We uncovered general preferences for information over reward across species at different timescales. On individual trials, tiles that were expected to be more informative tended to be selected sooner than rewarding ones for both humans and monkeys. Classic [41] and computational [14] reinforcement learning theories use rewards to assign credit to features in the environment. Many improvements to reinforcement learning have been proposed to deal with complexity in the environment, such as the introduction of options that group together many actions [42], the use of successor representations [43], or the use of belief distributions in partially observable environments to infer latent features from perceptible ones [44]. However, reward-driven reinforcement learning initially requires many samples over long periods of time to learn reward contingencies [45], especially if rewards are rarely delivered and the environment is constantly changing. Instead of reinforcement through the use of rewards, information in the form of changes in the identity and frequency of environmental features can be used to speed up learning by reducing uncertainty—formalizable in terms of information theory [46, 47]—about which features tend to occur together.

Humans are excellent information seekers [30, 48], efficiently using information to learn their environments [49, 50] and make inferences about the world in pursuit of their goals [51, 52, 53, 54, 55, 56]. Information-based approaches to learning either use that information directly in reinforcement-like processes [49, 50, 57, 58, 59, 60, 61] or indirectly in causal or structural inference [10, 62, 63, 64]. Since animals including humans often do not know the most informative choices, they must rely instead on hypothesis testing [65, 66, 67, 68, 69, 70, 71, 72, 73, 74], information foraging [36, 75, 76, 77, 78, 79, 80], or maximizing expected information gain [49, 50, 51, 81, 82, 83, 84, 85, 86] when making choices. Past studies have applied information-based approaches to a wide range of tasks, all with single choices and few latent

features [21, 51, 53, 55, 56, 61, 76, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97] (for review see [48]). Sequential information search is just beginning to be explored in relatively simple environments with one or two features [21, 56, 94, 98, 99, 100]. In addition, the recent focus in cognitive psychology on learning of latent features use tasks with environments that also often contain one feature and further fail to investigate learning over different timescales (within and across trials) [101, 102, 103], lack a decision-making component [104], utilize single choices during trials [105, 106, 107, 108, 109], or lack the choice complexity of the real world [110], though humans do adaptively explore for information in the hope of maximizing long-term rewards [111]. In our study, we extend this information seeking competence to temporally extended learning of many latent features by uncovering evidence for information-based foraging algorithms.

An unexpected and novel finding of our study was a signature of information foraging behavior that predicted the speed of learning latent features. Many trials in both humans and monkeys showed sequences of choices that reflected an area restricted search (ARS), persistent searching through a limited spatiotemporal region that is a hallmark of foraging behavior [34, 35, 112, 113, 114]. During learning, both humans and monkeys engaged in ARS for information. Decisions to search a different part of the grid were predicted by drops in information intake below the average across all choice outcomes. This average threshold rule is a signature of a standard computation during foraging: using the average outcome from choices across the environment to guide decisions to continue foraging locally or to leave the local area to look for new resources [38, 39, 40]. Importantly, humans that better matched the pattern of such foraging learned shapes quicker. Finally, in humans, this local search strategy was abandoned in favor of a reward-driven strategy once the shapes were learned. While exploratory, this finding can be used to ground predictions about information foraging in our and similar serial decision-making tasks.

Information search and foraging is important for theoretical approaches to complex cognition [36, 75, 112, 115]. The search for information forms the basis for understanding visual search [76, 116, 117] or chemotaxis [78, 118]. Information search also plays a role in explaining behavior on numerous tasks, including past studies on Battleship-like search tasks [49, 50, 119]. We extend these studies by reporting on the properties of sequences of choices to reveal evidence for a foraging information-intake threshold rule. Previous research [77, 120] has revealed indirect evidence for information foraging for humans surfing the internet. However, unlike our work reported here, these studies measured information in terms of associations between linguistic concepts [121]. In contrast, our findings are based on non-verbal search and are more generalizable to other effects reported in

the literature. The foraging framework more generally has been extended to numerous cognitive processes, including task-switching [122], internal word search [123], study time allocation [124], problem solving [125], memory search [126], semantic search [127], and even social interactions [128]. The foraging effects in these studies address reward-driven or performance-driven shifts between options. Our study reveals, for the first time, direct evidence for visuospatial information foraging in a cognitive task to learn latent features of the environment, extending the foraging framework to a new class of tasks.

We uncovered similar preferences for information in both humans and monkeys, though the extent to which animals other than humans search for and use information is less well characterized than it is for humans. Many animals including humans show a preference for advanced information about outcomes that cannot be used to adapt behavior (so-called 'observing responses' [129, 130, 131, 132], including pigeons [133], starlings [134], rats [135], monkeys [22, 23, 24, 26], and humans [25, 136, 137]). However, these studies tend to use environments with very few features and in the absence of information that can be used to help make later decisions. More recently, the preference for useful information, information that can be used to attain future rewards, has been studied in monkeys [27, 28, 29]. However, these studies focus on information gleaned on single trials and do not probe the cognitive capacity to learn multiple features over multiple timescales. Here we report the discovery that like humans, monkeys prefer to seek out useful information. Monkeys also 'over-explored' for information, persevering in information foraging even after shapes were learned. This difference may be due to humans' familiarity with game playing and shapes formed from blocks.

Our study has several 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 humans to monkeys, these findings remain to be generalized to other shape sets and other contexts, such as bigger grids. Second, our results crucially rely on various assumptions that can be challenged. For example, we used inferred changepoints as a measure of learning. Changepoints in the mean or variance of choices, however, might instead be the result of other processes such as attention, arousal, or boredom. We also assumed that participants' choice behavior can be understood as search in a restricted area (some contiguous set of tiles on the grid) and decisions to choose a non-neighboring tile reflect decisions to move to a different restricted area. Future experiments will need to manipulate confounding factors like attention, arousal, or boredom and test assumptions about restricted areas. 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 [138]. 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. In addition, fourth, we defined expected reward and expected information in terms of the outcomes from just the next choice; since information is used to maximize long-term reward rates, these two variables are confounded on our task in the long run. Finally fifth, 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.

While many studies have investigated how animals learn their environments, few have explored this learning in environments with multiple latent features using sequences of choices and across timescales. We discovered that, in a complex sequential decision-making task with many possible shapes, humans and monkeys were both driven more by information than reward. Evidence also showed that humans and monkeys foraged for information by sampling different areas of the grid in line with predictions from foraging theory. Finally and unexpectedly, the degree to which sequences of choices in humans matched foraging choice sequences predicted the speed at which the environment was learned. If confirmed by follow-up studies, this finding 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.

Methods

Our shape search task required participants to uncover shapes (composed of multiple tiles) in a 5x5 grid by selecting tiles that hide parts of the 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 (Fig. 1A; 1 shape ('H') occupied 7 tiles, and the others occupied 4 tiles) in one set. A trial refers to the sequences of choices required to finish revealing the shape. Humans and monkeys performed as many trials as possible per session. 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 participants, with a different location for each, and shapes overlapped at certain tiles. The shapes were selected on the basis of pre-existing data for the two monkeys that was collected while training them for a distinct electrophysiology study and were arbitrarily chosen by the experimenter (DLB). The set of five shapes did not change during the task, though participants were not instructed with regard to either the number of shapes or locations.

At trial start, participants 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 participants had unlimited time to choose a target. To select a tile, participants 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. Participants 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 shape search 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 participants chose tiles by mousing over the target until the choice registered and that tile was revealed. There were no training trials and we imposed no exclusion criteria on participants based on the number of trials completed. We collected data from 42 human participants (16m, 26f, average age 22y ± 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 reward maximizing choices on each trial. A reward maximizing choice is a choice of a tile that maximized the expected reward given what has already been revealed on the grid. For example, suppose that no tiles have been chosen yet at the start of the trial. The reward maximizing choice, then, 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 possible shapes. Then, if the two remaining shapes do not overlap at any further tiles, the choice of any tile that is part of either of the two shapes, but no other tiles, would be reward maximizing: given what has been revealed so far, only those tiles have any associated reward and they all have the same associated reward. For

each choice in each trial, the associated rewards for each remaining tile were calculated, and then the participants' actual responses compared to these calculations. The proportion of these reward maximizing choices was computed for every trial and plotted in Fig. 1B.

To compare this behavior to some baseline choice strategy, we simulated two agents. The first agent chose randomly, selecting any one of the remaining tiles with equal probability after each outcome. The second agent chose semi-randomly: if a hit had not yet been made on a trial, the tile selection was one of the remaining tiles with equal probability after each outcome; otherwise, if a hit had been made, any one of the adjoining remaining tiles was selected with equal probability. This semi-random strategy performs a local area search after getting the first hit. Because of the random nature of the choice, the algorithm can sometimes result in a dead-end, in which case it once again selects any of the remaining tiles with equal probability. A similar algorithm restricted to randomly choosing a city-block neighbor also performed poorly (not depicted). For humans, simulations were performed for the maximum number of trials across all participants (that is, the participant who performed the most trials was determined, and then the simulation performed for that number of trials; maximum number of trials = 109) (Fig. 1B, top panel), and for monkeys, for the same number of trials performed by 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 proportion of expected reward maximizing choices computed for each trial, and then these performances were averaged and the standard error of the mean (s.e.m.) computed and plotted (Fig. 1B, first agent: blue points and error bars; second agent: green points and error bars).

To quantify how quickly participants learned, we used a changepoint detection test on the mean and variance of choices for each separately [32, 33]. The changepoint detection test (cf. Gallistel 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, participants 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 participants 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. Inclan and Tiao 1994). Monkeys but not humans learned the shapes over multiple days. This variance was not calculated for those 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 (mean last changepoint trial for human: trial 45.72 ± 3.56; M1: trial 909; M2: trial 1191). This test failed to detect a changepoint for 3 of 42 human participants, who were removed from the learning analyses as a result.

A distribution of probabilities for each revealed shape, that is, the possible pattern that the grid could take once fully revealed, was used to compute information outcomes. These grids were formed by possible combinations of connected, filled tiles on a 3x3 grid that were then placed on the 5x5 grid. The shapes had to be at least 3 tiles large, no bigger than 8 tiles, 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 participants. While both monkeys had been trained on smaller grids 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 shape search task refers to the number of states.

The distribution of the probabilities of the states during trials, which we label '$\mathfrak{B}$', was updated 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 participant's uncertainty about the current trial's state. At the end of the trial, the Dirichlet distribution was updated by adding one 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 using 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.

To investigate the influence of expected rewards and expected 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 session, expected rewards for each tile, and expected information for each tile. Expected reward for a tile was defined as the number of times the tile proved rewarding divided by the number of times the tile was chosen. Expected information for a tile was defined as the mean change in entropy of $\mathfrak{B}$

if chosen and a hit was revealed or a miss was revealed. No interactions were included in the regression. The regression fits the following model to the data:

$$\log (p_{i<j} / p_{i \geq j}) = \beta_0 + \beta_1 * t_n + \beta_2 * EI + \beta_3 * ER$$

for choice numbers i, j, trial number t_n, expected information EI, and expected reward ER. In addition, we truncated the data to consider only those choices before the 10th choice in a trial; 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 grid. The results of this regression are plotted in Figure 2.

We analyzed pairs of choices by performing a mixed-effects binomial regression (*fitglme* in MATLAB). The dependent variable was choice of neighboring tile (1 = chose a neighbor, 0 = did not choose a neighbor). The first choice on each trial was removed for this regression (because there is no previous choice for comparison). The fixed-effect independent variables included trial number in session, choice number in trial, last choice information outcome, current choice expected information, last choice reward outcome, and current choice expected reward. The random-effect independent variable was subject identity. Fixed-effect independent variables were checked for correlation; pairwise R_{-}^2 values were all at or below 0.1385 except for the correlation between information outcomes and reward outcomes ($R_{-}^2 = 0.5593$). Significance was assessed against Bonferroni corrected p-values at 0.05. A similar binomial regression was run in the monkeys. Pairwise R_{-}^2 values were all at or below 0.0950 except for the correlation between choice number and expected information ($R_{-}^2 = 0.2998$) and information outcomes and reward outcomes ($R_{-}^2 = 0.4006$). All variance inflation factors for humans and monkeys for the main effect covariates were less than 2.51.

To plot sequences of choices, each choice on each trial was sorted according to whether the previous choice had been a neighboring tile. Those choices that were non-neighboring were labeled 'jumps'. 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 participants, 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 by 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. 4A) or reward (Fig. 4B) 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.

Participants' 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 [139, 140]. A classic foraging decision is 'patch leaving' where foragers must decide whether to continue foraging at a resource patch or to leave that patch to search for a new one. In the shape search 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 considered three features of choice sequences that reflect foraging inspired by [35, 38, 77]. First, while participants 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 the 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 $\underline{I}$, and subject s, let the forager score F_s be

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

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). The forager score was computed for every sequence of three choices in a row followed by a jump and averaged by subject. The final changepoint, which we used to quantify the end of learning, was then regressed against the average F_s (Fig. 5). As a comparison for this analysis, a similar score was constructed for rewards that used the reward outcomes following each choice in the sequences of three choices and the average reward outcomes across all choices.

Acknowledgements

The authors gratefully acknowledge the support of the National Institute for Health and National Institute for Drug Abuse (K99DA048748-01 to DLB) and the Presidential Scholars in Society and Neuroscience program at Columbia University (DLB).

References

1. Zaidi Q. Visual inferences of material changes: color as clue and distraction. *Wiley Interdisciplinary Reviews: Cognitive Science* **2**, 686-700 (2011).

2. Mnih V*, et al.* Human-level control through deep reinforcement learning. *Nature* **518**, 529-533 (2015).

3. Kaelbling LP, Littman ML, Cassandra AR. Planning and acting in partially observable stochastic domains. *Artificial intelligence* **101**, 99-134 (1998).

4. Maia TV. Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cognitive, Affective, & Behavioral Neuroscience* **9**, 343-364 (2009).

5. Braun DA, Mehring C, Wolpert DM. Structure learning in action. *Behavioural brain research* **206**, 157-165 (2010).

6. Gershman SJ, Blei DM, Niv Y. Context, learning, and extinction. *Psychol Rev* **117**, 197-209 (2010).

7. Wilson RC, Niv Y. Inferring relevance in a changing world. *Frontiers in human neuroscience* **5**, 189 (2012).

8. Dayan P, Berridge KC. Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience* **14**, 473-492 (2014).

9. Wilson RC, Takahashi YK, Schoenbaum G, Niv Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81**, 267-279 (2014).

10. Gershman SJ, Norman KA, Niv Y. Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences* **5**, 43-50 (2015).

11. Tervo DGR, Tenenbaum JB, Gershman SJ. Toward the neural implementation of structure learning. *Current opinion in neurobiology* **37**, 99-105 (2016).

12. Niv Y. Learning task-state representations. *Nature neuroscience* **22**, 1544-1553 (2019).

13. Salzman CD, Fusi S. Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annu Rev Neurosci* **33**, 173-202 (2010).

14. Sutton RS, Barto AG. *Reinforcement learning : an introduction*. MIT Press (1998).

15. Dayan P, Daw ND. Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience* **8**, 429-453 (2008).

16. Lee D, Seo H, Jung MW. Neural Basis of Reinforcement Learning and Decision Making. *Annual Review of Neuroscience* **35**, 287-308 (2012).

17. Tolman EC. Cognitive maps in rats and men. *Psychological review* **55**, 189 (1948).

18. Kolling N, Behrens TEJ, Mars RB, Rushworth MFS. Neural mechanisms of foraging. *Science* **336**, 95-98 (2012).

19. Kaplan R, King J, Koster R, Penny WD, Burgess N, Friston KJ. The neural representation of prospective choice during spatial planning and decisions. *PLoS biology* **15**, e1002588 (2017).

20. Kolling N, Scholl J, Chekroud A, Trier HA, Rushworth MF. Prospection, perseverance, and insight in sequential behavior. *Neuron* **99**, 1069-1082. e1067 (2018).

21. Meder B, Nelson JD, Jones M, Ruggeri A. Stepwise versus globally optimal search in children and adults. *Cognition* **191**, 103965 (2019).

22. Bromberg-Martin ES, Hikosaka O. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* **63**, 119-126 (2009).

23. Bromberg-Martin ES, Hikosaka O. Lateral habenula neurons signal errors in the prediction of reward information. *Nature neuroscience* **14**, 1209-1216 (2011).

24. Blanchard TC, Hayden BY, Bromberg-Martin ES. Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron* **85**, 602-614 (2015).

25. Iigaya K, Story GW, Kurth-Nelson Z, Dolan RJ, Dayan P. The modulation of savouring by prediction error and its effects on choice. *Elife* **5**, e13747 (2016).

26. Wang MZ, Hayden BY. Monkeys are curious about counterfactual outcomes. *Cognition* **189**, 1-10 (2019).

27. White JK*, et al.* A neural network for information seeking. *Nature communications* **10**, 1-19 (2019).

28. Foley NC, Kelly SP, Mhatre H, Lopes M, Gottlieb J. Parietal neurons encode expected gains in instrumental information. *Proceedings of the National Academy of Sciences* **114**, E3315-E3323 (2017).

29. Horan M, Daddaoua N, Gottlieb J. Parietal neurons encode information sampling based on decision uncertainty. *Nature neuroscience* **22**, 1327-1335 (2019).

30. Miller G. Informavores. In: *The study of information: Interdisciplinary messages* (eds Machlup F, Mansfield U). Wiley-Interscience (1983).

31. Osu R, Morishige K-i, Nakanishi J, Miyamoto H, Kawato M. Practice reduces task relevant variance modulation and forms nominal trajectory. *Scientific reports* **5**, 1-17 (2015).

32. Gallistel CR, Mark TA, King AP, Latham PE. 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* **27**, 354 (2001).

33. Inclan C, Tiao GC. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* **89**, 913-923 (1994).

34. Todd PM, Hills TT. Foraging in mind. *Current Directions in Psychological Science* **29**, 309-315 (2020).

35. Stephens DW, Krebs JR. *Foraging Theory*. Princeton University Press (1986).

36. Pirolli PLT. *Information foraging theory: Adaptive interaction with information*. Oxford University Press (2007).

37. Giraldeau L-A, Caraco T. Social foraging theory. In: *Social Foraging Theory*). Princeton University Press (2000).

38. Charnov EL. Optimal foraging, the marginal value theorem. *Theor Popul Biol* **9**, 129-136 (1976).

39. McNamara J. Optimal patch use in a stochastic environment. *Theoretical Population Biology* **21**, 269-288 (1982).

40. Davidson JD, El Hady A. Foraging as an evidence accumulation process. *PLoS computational biology* **15**, e1007060 (2019).

41. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical Conditioning II: Current Research and Theory* (eds Black AH, Prokasy WF). Appleton-Century-Crofts (1972).

42. Sutton RS, Precup D, Singh S. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence* **112**, 181-211 (1999).

43. Momennejad I. Learning Structures: Predictive Representations, Replay, and Generalization. *Current Opinion in Behavioral Sciences* **32**, 155-166 (2020).

44. Littman ML. A tutorial on partially observable Markov decision processes. *Journal of Mathematical Psychology* **53**, 119-125 (2009).

45. Kaelbling LP, Littman ML, Moore AW. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research* **4**, 237-285 (1996).

46. Shannon CE, Weaver W. *The Mathematical Theory of Communication*. University of Illinois Press (1963).

47. Crupi V, Nelson JD, Meder B, Cevolani G, Tentori K. Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search. *Cognitive Science* **42**, 1410-1456 (2018).

48. Coenen A, Nelson JD, Gureckis TM. Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin & Review* **26**, 1548-1587 (2019).

49. Gureckis T, Markant D. Active learning strategies in a spatial concept learning game. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*) (2009).

50. Markant D, Gureckis T. Does the utility of information influence sampling behavior? In: *Proceedings of the Annual Meeting of the Cognitive Science Society*) (2012).

51. Oaksford M, Chater N. A rational analysis of the selection task as optimal data selection. *Psychological Review* **101**, 608 (1994).

52. Oaksford M, Chater N. *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Psychology Press/Erlbaum (UK) Taylor & Francis (1998).

53. Nelson JD. Finding useful questions: on Bayesian diagnosticity, probability, impact, and information gain. *Psychological review* **112**, 979 (2005).

54. Oaksford M, Chater N. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press (2007).

55. Nelson JD, McKenzie CR, Cottrell GW, Sejnowski TJ. Experience matters: Information acquisition optimizes probability gain. *Psychological science* **21**, 960-969 (2010).

56. Nelson JD, Divjak B, Gudmundsdottir G, Martignon LF, Meder B. Children's sequential information search is sensitive to environmental probabilities. *Cognition* **130**, 74-80 (2014).

57. Schmidhuber J. Curious model-building control systems. In: *1991 IEEE International Joint Conference on Neural Networks.*). IEEE (1991).

58. Thrun S, Möller K. Active exploration in dynamic environments. In: *Advances in neural information processing systems*) (1992).

59. Thrun S. Exploration in active learning. *Handbook of Brain Science and Neural Networks*, 381-384 (1995).

60. Settles B. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* **6**, 1-114 (2012).

61. Markant DB, Gureckis TM. Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General* **143**, 94 (2014).

62. Griffiths TL, Tenenbaum JB. Structure and strength in causal induction. *Cognitive psychology* **51**, 334-384 (2005).

63. Kemp C, Tenenbaum JB. Structured statistical models of inductive reasoning. *Psychological review* **116**, 20 (2009).

64. Koechlin E. Prefrontal executive function and adaptive behavior in complex environments. *Current opinion in neurobiology* **37**, 1-6 (2016).

65. Wason PC. Reasoning. In: *New Horizons in Psychology* (ed Foss B) (1966).

66. Wason PC. Reasoning about a rule. *Quarterly journal of experimental psychology* **20**, 273-281 (1968).

67. Gregory R. On how little information controls so much behaviour. *Ergonomics* **13**, 25-35 (1970).

68. Snyder M, Swann WB. Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology* **36**, 1202 (1978).

69. Trope Y, Bassok M. Confirmatory and diagnosing strategies in social information gathering. *Journal of personality and social psychology* **43**, 22 (1982).

70. Klayman J, Ha Y-W. Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review* **94**, 211 (1987).

71. Siskind JM. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* **61**, 39-91 (1996).

72. Trope Y, Liberman A. Social hypothesis testing: Cognitive and motivational mechanisms. (1996).

73. Poletiek FH. *Hypothesis-testing behaviour*. Psychology Press (2013).

74. Markant DB, Settles B, Gureckis TM. Self-directed learning favors local, rather than global, uncertainty. *Cognitive science* **40**, 100-120 (2016).

75. Pirolli P, Card S. Information foraging. *Psychological review* **106**, 643 (1999).

76. Najemnik J, Geisler WS. Optimal eye movement strategies in visual search. *Nature* **434**, 387-391 (2005).

77. Fu W-T, Pirolli P. SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human–Computer Interaction* **22**, 355-412 (2007).

78. Vergassola M, Villermaux E, Shraiman BI. 'Infotaxis' as a strategy for searching without gradients. *Nature* **445**, 406 (2007).

79. Johnson A, Varberg Z, Benhardus J, Maahs A, Schrater P. The hippocampus and exploration: dynamically evolving behavior and neural representations. *Frontiers in human neuroscience* **6**, (2012).

80. Manohar SG, Husain M. Attention as foraging for information and value. *Frontiers in human neuroscience* **7**, 711 (2013).

81. Good IJ. Weight of evidence, corroboration, explanatory power, information and the utility of experiments. *Journal of the Royal Statistical Society: Series B (Methodological)* **22**, 319-331 (1960).

82. Myung JI, Pitt MA. Optimal experimental design for model discrimination. *Psychological review* **116**, 499 (2009).

83. Markant D, Gureckis T. Category learning through active sampling. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*) (2010).

84. Markant D, Gureckis T. Modeling information sampling over the course of learning. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*) (2011).

85. Tsividis P, Gershman S, Tenenbaum J, Schulz L. Information selection in noisy environments with large action spaces. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*) (2014).

86. Rich AS, Gureckis TM. Exploratory Choice Reflects the Future Value of Information. *Decision*, (2017).

87. Nelson J, Movellan J. Active inference in concept learning. *Advances in neural information processing systems* **13**, (2000).

88. Steyvers M, Tenenbaum JB, Wagenmakers EJ, Blum B. Inferring causal networks from observations and interventions. *Cognitive science* **27**, 453-489 (2003).

89. Schulz LE, Gopnik A, Glymour C. Preschool children learn about causal structure from conditional interventions. *Developmental science* **10**, 322-332 (2007).

90. Najemnik J, Geisler WS. Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision* **8**, 4 (2008).

91. Gopnik A. *The philosophical baby: What children's minds tell us about truth, love & the meaning of life*. Random House (2009).

92. Bonawitz EB*, et al.* Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. *Cognition* **115**, 104-117 (2010).

93. Cook C, Goodman ND, Schulz LE. Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition* **120**, 341-349 (2011).

94. Bramley NR, Lagnado DA, Speekenbrink M. Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **41**, 708 (2015).

95. Ruggeri A, Lombrozo T. Children adapt their questions to achieve efficient search. *Cognition* **143**, 203-216 (2015).

96. McCormack T, Bramley N, Frosch C, Patrick F, Lagnado D. Children's use of interventions to learn causal structure. *Journal of experimental child psychology* **141**, 1-22 (2016).

97. Rothe A, Lake BM, Gureckis TM. Do people ask good questions? *Computational Brain & Behavior* **1**, 69-89 (2018).

98. Meier KM, Blair MR. Waiting and weighting: Information sampling is a balance between efficiency and error-reduction. *Cognition* **126**, 319-325 (2013).

99. Yang SC-H, Lengyel M, Wolpert DM. Active sensing in the categorization of visual patterns. *Elife* **5**, e12215 (2016).

100. Nelson JD, Meder B, Jones M. Towards a theory of heuristic and optimal planning for sequential information search. (2018).

101. Badre D, Kayser AS, D'Esposito M. Frontal cortex and the discovery of abstract action rules. *Neuron* **66**, 315-326 (2010).

102. Wu CM, Schulz E, Speekenbrink M, Nelson JD, Meder B. Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour* **2**, 915-924 (2018).

103. Schulz E, Franklin NT, Gershman SJ. Finding structure in multi-armed bandits. *Cognitive psychology* **119**, 101261 (2020).

104. Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM. Neural representations of events arise from temporal community structure. *Nat Neurosci* **16**, 486-492 (2013).

105. Collins A, Koechlin E. Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol* **10**, e1001293 (2012).

106. Collins AG, Frank MJ. Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological review* **120**, 190 (2013).

107. Collins AG, Cavanagh JF, Frank MJ. Human EEG uncovers latent generalizable rule structure during learning. *J Neurosci* **34**, 4677-4685 (2014).

108. Donoso M, Collins AG, Koechlin E. Foundations of human reasoning in the prefrontal cortex. *Science* **344**, 1481-1486 (2014).

109. Collins AG. The cost of structure learning. *Journal of cognitive neuroscience* **29**, 1646-1655 (2017).

110. Xia L, Collins AG. Temporal and state abstractions for efficient learning, transfer, and composition in humans. *Psychological review*, (2021).

111. Wilson RC, Geana A, White JM, Ludvig EA, Cohen JD. Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General* **143**, 2074 (2014).

112. Hills TT. Animal Foraging and the Evolution of Goal-Directed Cognition. *Cognitive Science* **30**, 3-41 (2006).

113. Viswanathan GM, Da Luz MG, Raposo EP, Stanley HE. *The physics of foraging: an introduction to random searches and biological encounters*. Cambridge University Press (2011).

114. Hills TT, Kalff C, Wiener JM. Adaptive Lévy Processes and Area-Restricted Search in Human Foraging. *PloS one* **8**, e60488 (2013).

115. Hills TT, Todd PM, Goldstone RL. The central executive as a search process: priming exploration and exploitation across domains. *Journal of Experimental Psychology: General* **139**, 590 (2010).

116. Cain MS, Vul E, Clark K, Mitroff SR. A Bayesian optimal foraging model of human visual search. *Psychological Science*, 0956797612440460 (2012).

117. Wolfe JM. When is it time to move to the next raspberry bush? Foraging rules in human visual search. *Journal of vision* **13**, 1-17 (2013).

118. Calhoun AJ, Chalasani SH, Sharpee TO. Maximally informative foraging by Caenorhabditis elegans. *Elife* **3**, e04220 (2014).

119. Rothe A, Lake BM, Gureckis TM. Asking and evaluating natural language questions. In: *CogSci*) (2016).

120. Huberman BA, Pirolli PLT, Pitkow JE, Lukose RM. Strong regularities in world wide web surfing. *Science* **280**, 95-97 (1998).

121. Church K, Hanks P. Word association norms, mutual information, and lexicography. *Computational linguistics* **16**, 22-29 (1990).

122. Payne SJ, Duggan GB, Neth H. Discretionary task interleaving: heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General* **136**, 370 (2007).

123. Wilke A, Hutchinson J, Todd PM, Czienskowski U. Fishing for the right words: Decision rules for human foraging behavior in internal search tasks. *Cognitive Science* **33**, 497-529 (2009).

124. Metcalfe J, Jacobs WJ. People's study time allocation and its relation to animal foraging. *Behavioural processes* **83**, 213-221 (2010).

125. Payne S, Duggan G. Giving up problem solving. *Memory & Cognition* **39**, 902-913 (2011).

126. Hills TT, Jones MN, Todd PM. Optimal foraging in semantic memory. *Psychological review* **119**, 431 (2012).

127. Hills TT, Todd PM, Jones MN. Foraging in semantic fields: How we search through memory. *Topics in cognitive science* **7**, 513-534 (2015).

128. Turrin C, Fagan NA, Monte OD, Chang SWC. Social resource foraging is guided by the principles of the Marginal Value Theorem. *Scientific Reports* **7**, 11274 (2017).

129. Wyckoff Jr LB. The role of observing responses in discrimination learning. Part I. *Psychological review* **59**, 431 (1952).

130. Wyckoff L. Toward a quantitative theory of secondary reinforcement. *Psychological Review* **66**, 68 (1959).

131. Blanchard R. The effect of S− on observing behavior. *Learning and Motivation* **6**, 1-10 (1975).

132. Dinsmoor JA. Observing and conditioned reinforcement. *Behavioral and Brain Sciences* **6**, 693-704 (1983).

133. Roper KL, Zentall TR. Observing behavior in pigeons: The effect of reinforcement probability and response cost using a symmetrical choice procedure. *Learning and Motivation* **30**, 201-220 (1999).

134. Vasconcelos M, Monteiro T, Kacelnik A. Irrational choice and the value of information. *Scientific reports* **5**, 1-12 (2015).

135. Prokasy Jr WF. The acquisition of observing responses in the absence of differential external reinforcement. *Journal of comparative and physiological psychology* **49**, 131 (1956).

136. Kreps DM, Porteus EL. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica: journal of the Econometric Society*, 185-200 (1978).

137. Beierholm UR, Dayan P. Pavlovian-instrumental interaction in 'observing behavior'. *PLoS computational biology* **6**, e1000903 (2010).

138. Gottlieb J, Oudeyer P-Y. Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience* **19**, 758-770 (2018).

139. Calhoun AJ, Hayden BY. The foraging brain. *Current Opinion in Behavioral Sciences* **5**, 24-31 (2015).

140. Barack DL, Platt ML. Engaging and Exploring: Cortical Circuits for Adaptive Foraging Decisions. In: *Impulsivity*). Springer (2017).