

Normative theory of patch foraging decisions

Zachary P Kilpatrick*

*Department of Applied Mathematics, University of Colorado, Boulder, Colorado 80309, USA and
Department of Physiology & Biophysics, University of Colorado School of Medicine, Aurora, Colorado, USA*

Jacob D Davidson†

*Department of Collective Behavior, Max Planck Institute of Animal Behavior, Konstanz, Germany
Department of Biology, University of Konstanz, Konstanz, Germany and
Centre for the Advanced Study of Collective Behavior, University of Konstanz, Konstanz, Germany*

Ahmed El Hady‡

Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, USA

(Dated: April 22, 2020)

Foraging is a fundamental behavior as animals' search for food is crucial for their survival. Patch leaving is a canonical foraging behavior, but classic theoretical conceptions of patch leaving decisions lack some key naturalistic details. Optimal foraging theory provides general rules for when an animal should leave a patch, but does not provide mechanistic insights about how those rules change with the structure of the environment. Such a mechanistic framework would aid in designing quantitative experiments to unravel behavioral and neural underpinnings of foraging. To address these shortcomings, we develop a normative theory of patch foraging decisions. Using a Bayesian approach, we treat patch leaving behavior as a statistical inference problem. We derive the animals' optimal decision strategies in both non-depleting and depleting environments. A majority of these cases can be analyzed explicitly using methods from stochastic processes. Our behavioral predictions are expressed in terms of the optimal patch residence time and the decision rule by which an animal departs a patch. We also extend our theory to a hierarchical model in which the forager learns the environmental food resource distribution. The quantitative framework we develop will therefore help experimenters move from analyzing trial based behavior to continuous behavior without the loss of quantitative rigor. Our theoretical framework both extends optimal foraging theory and motivates a variety of behavioral and neuroscientific experiments investigating patch foraging behavior.

CONTENTS

		A. Homogeneous environment	17
		B. Binary environments	19
I. Introduction	1	VI. Discussion	20
II. Sequential updating foraging model	3	A. Summary	20
A. Full model	3	B. Relation to other work	20
B. Idealizing limits	3	C. Experimental applications: Behavior and neuroscience	23
III. Minimizing time to find high patch in non-depleting environments	5	D. Model extensions	24
A. Binary environment	5	Acknowledgments	24
B. Three patch types	8	References	24
C. More patch types	9		
D. Continuum limit	10		
IV. Depletion- vs. uncertainty-driven decisions in depleting environments	11		
A. Homogeneous environments	11		
B. Binary environments	12		
C. Returning to a previously depleted patch	15		
V. Learning the distribution of patch types	17		

I. INTRODUCTION

Nearly all animals forage, as it is essential to acquire energy for survival through efficient search and harvesting of food resources. Foraging behavior is performed by small organisms like *C. Elegans* [1, 2] and *Drosophila* [3, 4], animals with large spatial ranges like birds [5, 6], and mammals like rodents [7, 8], monkeys [9, 10], and humans [11, 12]. In addition to its universality across animal species, foraging engages multiple cognitive computations such as learning of food distributions across spatiotemporal scales, statistical inference of food availability, route planning and decision-

* zpkipat@colorado.edu

† j davidson@ab.mpg.de

‡ ahady@princeton.edu

making [11]. As such, foraging utilizes multiple neural systems in the brain [13], offering researchers the opportunity to probe these neural computations simultaneously. In addition to studying the proximal neural and behavioral mechanisms shaping foraging across species, one can also ask how these processes and their interplay have been shaped by natural selection to optimize returns in the face of environmental and physiological constraints [11, 14–16] opening up the opportunity to perform evolutionary quantitative behavioral studies. For the aforementioned reasons, there has been an increased interest in studying foraging in a neuroscience context [13, 17–19].

Since the space of foraging behaviors is vast, it is crucial to build a conceptual framework starting with a tractable behavior that is sufficiently flexible and rich. Adhering to this aim, patch foraging is both a common and understandable behavior, described as follows [20, 21]: an animal enters a patch of food, harvests resources, and then leaves to search for another patch of food [12]. An animal’s behavior can thus be quantified by its patch residence time distribution, travel time distribution, the amount of food resources consumed, and the movement pattern in-between patches. Moreover, the animal’s reward rate can be computed by its food intake divided by time. A basic result in behavioral ecology, the marginal value theorem (MVT) states that an animal can optimize energy intake by leaving its current patch when the current reward rate falls below the global average reward rate of the environment [22]. This gives a theoretical foundation for assessing optimal decision-making, which has also been validated in many behavioral studies [20, 21, 23–30].

However, the MVT does not describe mechanistically how an animal uses its foraging experience to learn key environmental features like the distribution of food availability, and assumes the animal already knows the overall average reward rate of the environment. Although mechanistic models of foraging have been proposed [20, 21, 26, 29, 31], these were not derived based on principles of statistical inference, and therefore cannot address how optimal decisions depend on the resource environment and uncertainty in the statistics of food availability.

In behavioral ecology, many studies have considered a Bayesian approach to patch leaving, asking how an animal uses available information to make optimal foraging decisions [32–49]. However, because most studies consider a specific or narrow range of environmental conditions, it is not clear how a Bayesian approach relates to other mechanistic models of foraging decisions, and how such an approach may assist in deriving possible neural implementations. To date there is not a complete normative theory of how animals make patch leaving decisions under a broad class of environmental conditions.

The aim of our study is thus to develop a Bayesian framework of patch leaving behavior by treating decisions as a statistical inference problem. Our framework con-

nects normative theory of foraging decisions with mechanistic drift-diffusion models, as proposed in [31].

To realize our theoretical framework, we perform the following model derivations and analysis:

- We consider a series of idealizing limits such as non-depleting patches, homogeneous environments, and binary environments. Using probabilistic sequential updating, we derive stochastic differential equations (SDEs) for the belief about the arrival rate of food in the current patch. This generates analytically tractable models associated with optimal patch leaving strategies.
- We consider a series of more naturalistic conditions such as depleting patches, environments in which the forager returns to depleted patches, and environments with many or a continuum of patch types.
- We derive optimal decision strategies for patch leaving that either maximize the long term reward rate or minimize the time to enter and remain in a high yielding patch across all the aforementioned environmental structures.
- We study how animals can learn the distribution of patch types in both non-depleting and depleting environments. This allows us to study how the rate at which the environmental distribution is learned depends on environmental parameters.

Following the aforementioned modeling approach, we arrive at a number of general conclusions concerning the best strategies for foragers to adopt in different environments:

- In idealizing limits, patch leaving decision statistics are described by solutions of first passage time problems for SDEs with absorbing boundaries.
- In non-depleting environments, in which consumption does not diminish food availability, an ideal forager should minimize their time to find and remain in the highest yielding patch in the environment. This time decreases both as high yielding patches become more common and as the highest yielding patch becomes more discriminable.
- In depleting environments, an ideal forager must estimate the current arrival rate of food within a patch and when this becomes too low, depart the patch.
- Across a wide range of task parameters, the forager should leave a patch when the current reward rate is matched to the environmental reward rate, as in the MVT. However, if there is a high level of uncertainty about the food availability within a patch, even an ideal forager will tend to stay too long in low yielding patches (overharvesting) and not long enough in high yielding patches (underharvesting)

- When the environment is fully depleting, such that a forager may return again to patches they may have already partially depleted, they can minimize the time to deplete the environment by aiming to fully deplete patches before leaving them.

Our aforementioned results highlight that we have found several key deviations from the typical MVT results of classic optimal foraging theory. These deviations arise due to either limited information available to the forager, slow patch depletion, or full environmental depletion.

Another key aspect of our model that extends beyond classic optimal foraging theory is the ability of our ideal forager to learn the underlying food arrival rates of the environment. The forager learns patch arrival rates more rapidly when rates are high, since food encounters provide more information than times between food encounters. Also, arrival rates are learned more quickly in depleting environments than in non-depleting environments, since successive food encounters rule out low initial arrival rates. Combined with the broad range of optimal foraging strategies derived assuming the observer has full knowledge of its environment, our analysis provides a number of touchstones to be compared with patch leaving statistics from future and past behavioral studies.

We anticipate that our theoretical framework will provide a platform to study behavioral and neural mechanisms of naturalistic decision-making in a similar manner as trained decision-making behavior is studied within systems neuroscience [17]. Our model is therefore useful in opening up the space for laboratory experiments that mimic naturalistic patch foraging dynamics and for generating testable hypothesis under different ecologically relevant experimental designs.

II. SEQUENTIAL UPDATING FORAGING MODEL

We model an animal searching a large arena representing a natural environment with distributed patches of food (Fig. 1A). When the animal enters a food patch, it consumes food within the patch until it decides to leave for another patch. The parameters characterizing the behavior are: environmental distribution of initial food density $p(\lambda_0)$, patch size A , food chunk size c , and mean travel time between patches τ . In this work, we assume the animal knows the patch size, chunk size, and the travel time (which is held fixed). Note, we could model the learning of these parameters as separate sequential updating processes (See Discussion for more details). Our interest is in how an observer can learn the food density within and across patches over time and use this information to guide an efficient foraging strategy.

A. Full model

Under simplifying assumptions, we derive a sequential sampling model for an ideal observer's posterior of their current patch's food yield rate $\lambda(t)$. To model the random timing of food encounters, we assume they are Poisson-distributed with a rate $\lambda(t) = \lambda_0 - \rho K(t)$ that decreases with $K(t)$ the number of food chunks found so far, and ρ is the impact of each food chunk on the arrival rate. Thus, the time t_K between the K th and $(K + 1)$ th encounter is drawn from the exponential distribution $(\lambda_0 - \rho K)e^{-(\lambda_0 - \rho K)t}$. We assume for now that the observer knows and initializes their belief with the prior $p(\lambda_0)$ when arriving in a new patch (Fig. 1B). In Section V, we extend our model to consider the process by which the distribution $p(\lambda_0)$ is learned. Given the food encounter sequence $x(t) = K'(t) = \sum_{j=1}^{K(t)} \delta(t - t_j)$, an ideal observer updates their belief about the current arrival rate λ according to

$$p(\lambda|x(t)) = p(x(t)|\lambda) \frac{p_0(\lambda + K(t)\rho)}{p(x(t))} \propto \frac{(\lambda/\rho + K)!}{(\lambda/\rho)!} e^{-\lambda t} p_0(\lambda + K\rho), \quad \lambda \geq 0. \quad (1)$$

Food encounters shape the observer's belief about the current yield rate $\lambda(t)$ in two main ways: (1) they give evidence of higher yield rates, since encounters are more probable in high yielding patches; and (2) they deplete the patch, decrementing the yield rate, as described by the ρ terms in the posterior update. Thus, there is a tension between the depleting effects of food encounters, and the evidence they provide for higher yield rates (Fig. 1B).

To better understand how an ideal observer's evidence accumulation process evolves, we consider a few different idealizing limits of Eq. (1). Analyzing these reveals distinct patterns in the belief dynamics as determined by task parameters, as we will show.

B. Idealizing limits

Non-depletion. In the limit $\rho \rightarrow 0^+$, food encounters do not deplete the rate of food arrival, which could be due to slow search rates or large patch areas [31]. Taking this limit, Eq. (1) simplifies considerably to

$$p(\lambda|x(t)) = \frac{p_0(\lambda)}{p(t_{1:K(t)})} \lambda^{K(t)} e^{-\lambda t}, \quad \lambda \geq 0,$$

so the log-likelihood update for $\lambda \geq 0$ becomes

$$\log p(\lambda|x(t)) = K(t) \log \lambda - \lambda t - \log p(t_{1:K(t)}) + \log p_0(\lambda), \quad (2)$$

showing that food encounters provide a pulse of evidence that increases with λ , and a lack of food results in a linear decrease of the log likelihood of the arrival rate λ .

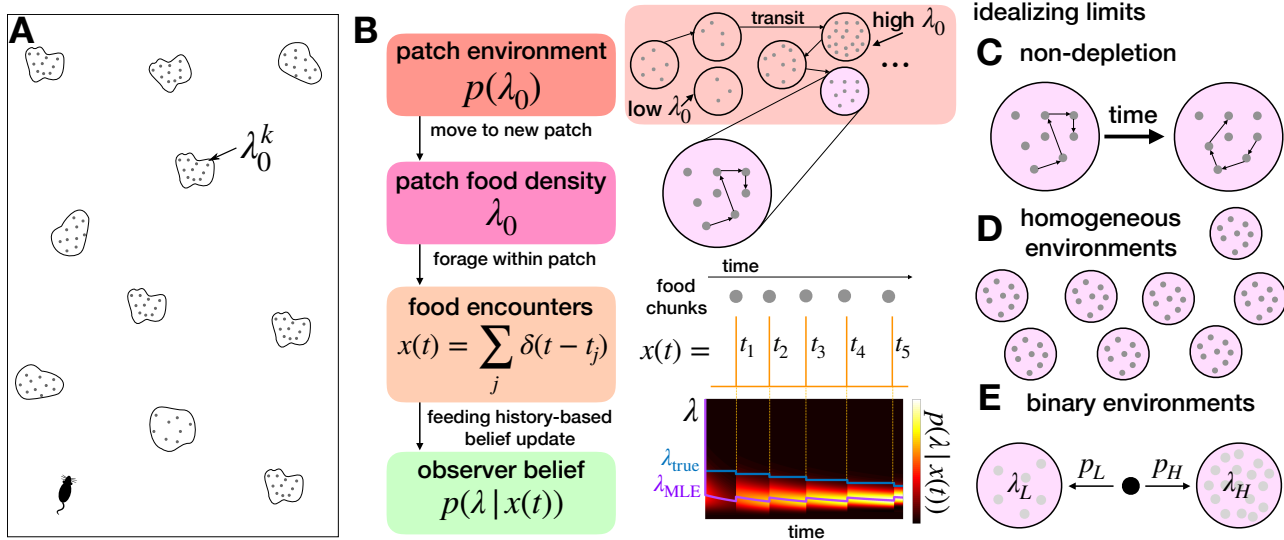


FIG. 1. Patch foraging task and model. **A.** Task environment: An animal enters an arena and forages from patches, each patch k has some (possibly distinct) initial food yield rate λ_0^k . **B.** Ideal observer foraging model: Initial patch yield is drawn from the distribution $p(\lambda_0)$, generating random food encounter times $t_{1:K}$, and updating belief of current food arrival rate $\lambda(t)$ for patch. The maximum likelihood estimate λ_{MLE} gets approaches the true λ_{true} arrival rate over time. *Model idealizations:* **C.** Non-depletion: Assuming patches are depleted slowly enough to be approximated as ‘non-depleting.’ **D.** Homogeneous environments: All patches have the same initial yield rate λ_0 . **E.** Binary environment: Two possible initial patch yield rates $\lambda_H > \lambda_L$ occur with probability p_H and p_L .

In binary environments, with two possible arrival rates $\lambda_H > \lambda_L$, we have $p_0(\lambda) = p_H \delta(\lambda - \lambda_H) + p_L \delta(\lambda - \lambda_L)$ where $p_H + p_L = 1$. Eq. (2) collapses to a scalar update equation for the log-likelihood ratio (LLR),

$$y(t) \equiv \log \frac{p(\lambda_H | x(t))}{p(\lambda_L | x(t))} = \underbrace{K(t) \log \frac{\lambda_H}{\lambda_L}}_{\text{food encounters}} - \underbrace{(\lambda_H - \lambda_L)t}_{\text{lack of food}} + \underbrace{\log \frac{p_H}{p_L}}_{\text{prior}},$$

which can be written equivalently as a stochastic differential equation (SDE) for the rate-of-change of the LLR:

$$\frac{dy}{dt} = \underbrace{\log \frac{\lambda_H}{\lambda_L} \sum_{j=1}^{\infty} \delta(t - t_j)}_{\text{food encounters}} - \underbrace{(\lambda_H - \lambda_L)}_{\text{lack of food}}, \quad (3)$$

with initial condition set by the prior $y(0) = \log \frac{p_H}{1 - p_H}$. Eq. (3) has a form similar to classic evidence accumulation models commonly used to model psychophysical data from decision-making tasks [50, 51], as well as to previous models of foraging decisions [31]. Due to its simple form, it can be analyzed explicitly to determine how long term statistics are shaped by task parameters.

Homogeneous environments. Another way of simplifying Eq. (1) is to consider homogeneous environments where knowledge of the patch yield rate is perfect. To see this, note that if $p_0(\lambda) = \delta(\lambda - \lambda_0)$, then Eq. (1) implies $p(\lambda | x(t)) = \delta(\lambda - (\lambda_0 - K(t)\rho))$. In other words, the observer has perfect knowledge that $\lambda(t) = \lambda_0 - K(t)\rho$.

Binary depleting environments. We also explore belief updating and patch departure strategies in depleting binary environments, revealing two phases of yield rate inference: an initial discrimination phase followed by depletion based decrementing of the yield rate estimate. In this case, we have $p_0(\lambda) = p_H \delta(\lambda - \lambda_H) + p_L \delta(\lambda - \lambda_L)$, and the associated LLR has a rate-of-change given by the non-autonomous SDE:

$$\frac{dy}{dt} = \sum_{j=1}^{K_{\max}} \log \frac{\lambda_H - (j-1)\rho}{\lambda_L - (j-1)\rho} \cdot \delta(t - t_j) - (\lambda_H - \lambda_L), \quad (4)$$

where $y(0) = \log \frac{p_H}{1 - p_H}$ as in the non-depleting case.

These idealizing limits all afford some level of tractability. However, we first derive optimal patch leaving strategies in non-depleting binary environments. In this case we can explicitly derive first passage time statistics associated with patch departures, allowing us to examine how efficient patch leaving depends on environmental parameters like patch discriminability (λ_H/λ_L) and high patch prevalence (p_H). The trends we find in this idealized case help us develop and interpret optimal strategies in more complex cases like depleting environments and in environments with more than two patch types.

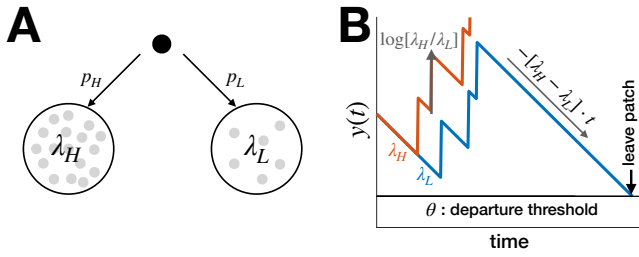


FIG. 2. **Belief updating in binary non-depleting environments.** **A.** Distribution of high (λ_H) and low (λ_L) patch types. **B.** Patch type belief $y(t) = \log \frac{p(\lambda_H|x(t))}{p(\lambda_L|x(t))}$ increases with food encounters and decreases between until $y(t) = \theta$ and the observer departs the patch.

III. MINIMIZING TIME TO FIND HIGH PATCH IN NON-DEPLETING ENVIRONMENTS

A forager in an environment with non-depleting patches is best off locating a patch with the highest yield and remaining there. In binary environments with two yield rates $\lambda_H > \lambda_L$ (Fig. 2A) and belief update given by Eq. (3), we show an efficient patch departure strategy is to set a single threshold on the LLR (or likelihood) so that when the forager is sufficiently confident they are in a low yielding patch, they leave. We extend this approach to consider environments with multiple patch types (e.g., three or a continuum) and show a similar thresholding strategy can be used for the forager to most quickly find and remain in the highest yielding patches.

A. Binary environment

Considering Eq. (3), we analyze a formula for the long term reward rate of the forager, and identify optimal parameters for an LLR threshold crossing patch leaving strategy (Fig. 2B). The reward rate is maximized in the long time limit as long as the forager eventually ends up in a high yielding patch, which occurs via a LLR thresholding strategy. A more thorough analysis of the optimal strategy using dynamic programming could be performed, but as the system does not fit the conditions of decision processes that require time-varying thresholds, we expect this would yield the same result [52, 53].

Energy intake rate. To begin the analysis, define the reward rate function to be maximized over a particular patch leaving policy π for a given total foraging time T :

$$R^\pi(T) := \frac{E^\pi(T)}{T} = \frac{\langle E_H^\pi(T_H, T) \rangle + \langle E_L^\pi(T_L, T) \rangle}{T},$$

which is the sum of mean food consumed from high and low yielding patch visits $\langle E_{H,L}^\pi(T_L, T) \rangle$ divided by the total foraging time T . The quantity $E^\pi(T)$ is the average number of food chunks consumed. Since policies can condition patch leaving on food chunks consumed, the average food chunk count for a finite time t in a high/low

patch will likely not be the mean of the Poisson arrival process $\lambda_{H,L}t$. However, consider the limit $T \rightarrow \infty$ of long foraging time:

$$\lim_{T \rightarrow \infty} R^\pi(T) = R_H^\pi(\infty) + R_L^\pi(\infty),$$

in which there are three possible scenarios: (a) the forager spends a nonzero fraction of time in both patches so both $R_{H,L}^\pi(\infty) \neq 0$; (b) the forager only spends a nonzero fraction of time in the high yielding patch so $R_L^\pi(\infty) \equiv 0$; or (c) the forager only spends a nonzero fraction of time in the low yielding patch so $R_H^\pi(\infty) \equiv 0$.

In policies $\pi(a)$ leading to case (a), the forager will spend some fraction of time transitioning between patches and the rest of the time in patches. Define r_H and r_L to be the fraction of time spent in the high and low yielding patches ($r_H + r_L < 1$ due to transits), then

$$\lim_{T \rightarrow \infty} R^{\pi(a)}(T) = r_H \lambda_H + r_L \lambda_L$$

for such policies, as the average yield rates $\lambda_{H,L}$ will be recovered in each patch in the long time limit.

On the other hand, any policy which leads to the forager only visiting one patch type for a finite amount of time (a nonzero time fraction as $T \rightarrow \infty$) must involve the forager remaining in one patch indefinitely. Otherwise, if they continued patch leaving indefinitely, they would spend a nonzero fraction of time in patches of the other type. As such, policies $\pi(b)$ and $\pi(c)$ causing the forager to (b) stay in the high patch and (c) stay in the low patch imply

$$\lim_{T \rightarrow \infty} R^{\pi(b)}(T) = \lambda_H \quad \text{and} \quad \lim_{T \rightarrow \infty} R^{\pi(c)}(T) = \lambda_L$$

As the forager's fraction of time spent in the high (low) patch type converges to unity, $r_L \rightarrow 0$ ($r_H \rightarrow 0$) in case b (c). Clearly, all policies of type $\pi(b)$ will maximize long term energy intake rate.

Thus, we can maximize $\lim_{T \rightarrow \infty} R^\pi(T)$ via any policy whereby the forager finds and remains in a high yielding patch. To do this, we set a lower threshold $\theta < \log \frac{p_H}{1 - p_H}$ on the belief $y(t)$ in Eq. (3), so the forager leaves given sufficient evidence that the patch is low yielding. We will show subsequently that this leads to a nonzero probability of remaining in a high yielding patch indefinitely each time one arrives there. Thus, there is complete certainty the forager will eventually find and remain in the high patch for all time.

Minimizing time to arrive in high yielding patch. Given a constant LLR thresholding strategy, we can derive an implicit equation for how the mean time to arrive and remain in the high patch $\bar{T}_{\text{arrive}}(\theta)$ depends on θ . This quantity can be computed from the high patch escape probability $\pi_H(\theta)$, the high patch mean visit time $\bar{T}_H(\theta)$ when departing, the low patch mean visit time $\bar{T}_L(\theta)$, the known patch fraction p_H , and mean transit time τ . Patch departure time statistics can be determined explicitly by solving the mean first passage time

problem for the SDE in Eq (3) with absorbing boundary at $y = \theta$. With these quantities in hand, we can compute

$$\begin{aligned} \bar{T}_{\text{arrive}}(\theta) = & \frac{\pi_H(\theta)}{1 - \pi_H(\theta)} \left(\bar{T}_H(\theta) + \tau + \frac{1 - p_H}{p_H} (\bar{T}_L(\theta) + \tau) \right) \\ & + \frac{1 - p_H}{p_H} (\bar{T}_L(\theta) + \tau), \end{aligned} \quad (5)$$

where the first term accounts for the number of high patch visits required to remain in the high yielding patch and the second term accounts for the time needed to leave the low yielding patch when starting there. For thresholds further from $y(0) = \log \frac{p_H}{1 - p_H}$, the escape probability $\pi_H(\theta)$ decreases and mean exit times $\bar{T}_{H,L}(\theta)$ increase. Intuitively, if $\theta = \log \frac{p_H}{1 - p_H}$, departure from each patch is immediate, whereas for $\theta \rightarrow -\infty$ then $\pi_H(-\infty) \rightarrow 0$ but also $\bar{T}_{H,L}(\theta) \rightarrow \infty$ since the forager never obtains sufficient evidence to leave a patch. Thus, we expect an interior solution θ to $\text{argmin}_{\theta \in (-\infty, \log \frac{p_H}{1 - p_H})} [\bar{T}_{\text{arrive}}(\theta)]$. We now compute the components of Eq. (5) using first passage time methods for SDEs [54].

Patch departure time statistics. The problem of finding the time for a forager to leave a patch can be formulated as a first passage time calculation. To obtain explicit results, we derive the corresponding backward Kolmogorov equation of Eq. (3). To do so, first note that the forward Kolmogorov equation is

$$p_t(y, t) = +ap_y(y, t) + \lambda \Psi(p(y, t), y) - \lambda p(y, t) \equiv \mathcal{L}p(y, t)$$

where $a = \lambda_H - \lambda_L$ and $\lambda \in \{\lambda_H, \lambda_L\}$ is the arrival rate of the current patch, and

$$\Psi(p(y, t), y) = \begin{cases} p(y - b, t), & y > \theta + b, \\ 0, & \theta < y < \theta + b, \end{cases}$$

accounts for the fact that jumps in the belief y due to food encounters can only lead to beliefs that are at least one jump length $b = \log \frac{\lambda_H}{\lambda_L}$ from threshold θ . Patch departures are represented by an absorbing boundary condition at the threshold: $p(\theta, t) = 0$. Defining the L^2 inner product $\langle u, v \rangle = \int_{\theta}^{\infty} u(y)v(y)dy$ over real functions, we determine the adjoint \mathcal{L}^* as $\langle \mathcal{L}u, v \rangle = \langle u, \mathcal{L}^*v \rangle$, yielding the right hand side of the corresponding backward Kolmogorov equation [54]:

$$\begin{aligned} q_t = & -aq_y(y', t|y, 0) + \lambda q(y', t|y + b, 0) - \lambda q(y', t|y, 0) \\ \equiv & \mathcal{L}^*q(y', t|y, 0), \end{aligned} \quad (6)$$

with associated initial and boundary conditions $p(y', 0|y, 0) = \delta(y' - y)$ for $y, y' > \theta$ and $p(y', t|\theta, 0) = 0$ for $t > 0$.

Now to obtain patch departure statistics, we can integrate Eq. (6) over all possible belief values $y' \in (\theta, \infty)$ to determine the probability that the patch departure time T is after the time t . We define the survival function conditioned on the starting belief y as

$$P(T > t|y) = \int_{\theta}^{\infty} q(y', t|y, 0)dy' \equiv G^j(y, t),$$

where $j \in \{H, L\}$ indexes the true patch type λ_j . Upon integrating Eq. (6), we find

$$G_t^j = -aG_y(y, t) + \lambda_j [G(y + b, t) - G(y, t)] \equiv \mathcal{L}^*G^j, \quad (7)$$

on $y \in (\theta, \infty)$, and note that the absorbing boundary condition ensures $G^j(\theta, t) = 0$ for $t > 0$.

Moreover, by taking the limit $t \rightarrow 0$ of Eq. (7), we can derive an equation for the escape probability $\pi_j(y) \equiv G^j(y, 0) = P(T \geq 0|y)$, which is the probability the belief ever crosses threshold given it begins at y and $t = 0$:

$$0 = -a\pi'(y) + \lambda_j [\pi(y + b) - \pi(y)] \equiv \mathcal{L}^*\pi_j(y), \quad (8)$$

with absorbing boundary condition $\pi_j(\theta) = 1$. Eigen-solutions of Eq. (8) take the form $\pi_j(y) = Ae^{ry}$ with corresponding transcendental characteristic equation

$$\begin{aligned} 0 = & -ar + \lambda_j e^{rb} - \lambda_j = -(\lambda_H - \lambda_L)r + \lambda_j \left(\frac{\lambda_H}{\lambda_L} \right)^r - \lambda_j \\ \equiv & C_j(r). \end{aligned} \quad (9)$$

As $\lambda_H > \lambda_L$, $C_j''(r) = \lambda_j b^2 (\lambda_H/\lambda_L)^{r-2} > 0$, this implies $C_j(r)$ is concave up and so has at most two roots: one is $r_1 = 0$, and when $j = H$ ($j = L$) the other root is $r_2 = -1$ ($r_2 = 1$).

To superpose solutions in each case ($j \in \{H, L\}$), note when $j = L$, the average drift

$$\langle \dot{y} \rangle = \lambda_L \log \frac{\lambda_H}{\lambda_L} - (\lambda_H - \lambda_L) < 0$$

of the belief in Eq. (3) is down toward threshold and we expect $\lim_{y \rightarrow \infty} \pi_L(y) = 1$. Applying this and the boundary condition $\pi_L(\theta) = 1$, we see $\pi_L(y) \equiv 1$ when $\lambda = \lambda_L$. However, when $\lambda = \lambda_H$, the average drift

$$\langle \dot{y} \rangle = \lambda_H \log \frac{\lambda_H}{\lambda_L} - (\lambda_H - \lambda_L) > 0$$

is away from threshold, implying $\lim_{y \rightarrow \infty} \pi_H(y) = 0$ and $\pi_H(\theta) = 1$, yielding $\pi_H(y) = e^{\theta - y}$.

Now to compute the mean first passage time (MFPT) in either case, note that the probability density function (pdf) $f_j(t)$ of exit times in case j can be calculated from the normalized survival function

$$\lim_{\Delta t \rightarrow 0} \frac{P(T > t|y) - P(T > t + \Delta t|y)}{\Delta t \cdot P(T \geq 0|y)} = -\frac{G_t^j(y, t)}{\pi_j(y)} \equiv f_j(t).$$

The MFPT $\bar{T}_j(y)$ is then obtained by integrating against the pdf:

$$\bar{T}_j(y) = \int_0^{\infty} t f_j(t) dt = -\frac{\int_0^{\infty} t G_t^j(y, t) dt}{\pi_j(y)} = \frac{\int_0^{\infty} G^j(y, t) dt}{\pi_j(y)}.$$

Integrating Eq. (7) over $t \in (0, \infty)$ and defining $\mathcal{T}_j(y) := \int_0^{\infty} G^j(y, t) dt = \pi_j(y) \bar{T}_j(y)$, we see

$$-\pi_j(y) = -a\mathcal{T}_j'(y) + \lambda_j [\mathcal{T}_j(y + b) - \mathcal{T}_j(y)] \equiv \mathcal{L}^*\mathcal{T}_j. \quad (10)$$

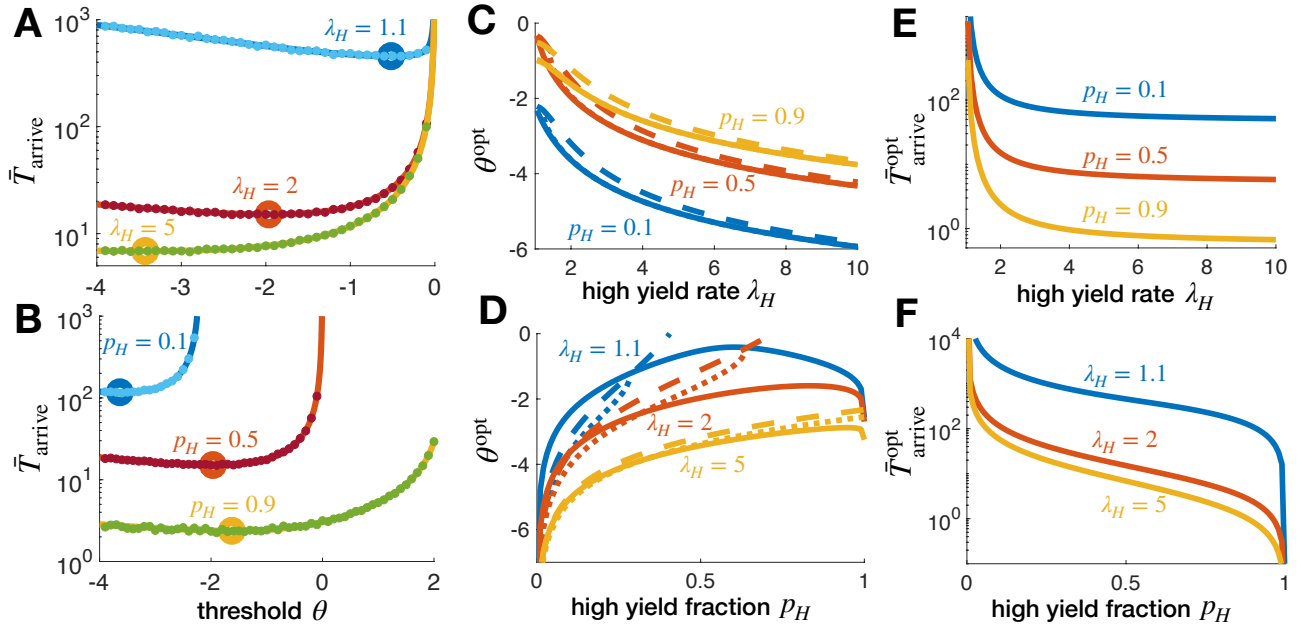


FIG. 3. Statistics of high patch identification in binary non-depleting environments. **A,B.** The mean time to arrive and remain in a high yield patch varies nonmonotonically with departure threshold θ and decreases as the patch discriminability λ_H/λ_L and high yield fraction p_H are increased. Solid lines are the analytical result in Eq. (5) and dots are averages over 10^4 Monte Carlo simulations. $\lambda_H = 2$ and $p_H = 0.5$ are fixed unless indicated. **C,D.** Departure threshold θ^{opt} minimizing the time to arrive in the high patch decreases with λ_H/λ_L and varies nonmonotonically with p_H . Solid lines are obtained numerically by solving Eq. (11), dotted lines are the Lambert W function approximation (Eq. (12)), and dashed lines are the asymptotic expansion in logarithms (Eq. (13)). **E,F.** The minimal mean time $\bar{T}_{\text{arrive}}^{\text{opt}}$ to arrive in a high yield patch decreases with λ_H/λ_L and p_H . We fix $\lambda_L = 1$ and $\tau = 5$.

For $j = L$, the boundary conditions $\bar{T}_L(\theta) = 0$, $\pi_L(\theta) = 1$, $\lim_{y \rightarrow \infty} \bar{T}_L(y) = \infty$, and $\lim_{y \rightarrow \infty} \pi_L(y) = 1$ imply

$$\bar{T}_L(y) = \frac{\mathcal{T}_L(y)}{\pi_L(y)} = \frac{y - \theta}{(\lambda_H - \lambda_L) - \lambda_L \log \frac{\lambda_H}{\lambda_L}}.$$

For $j = H$, the boundary conditions $\bar{T}_H(\theta) = 0$, $\pi_H(\theta) = 1$, $\lim_{y \rightarrow \infty} \bar{T}_H(y) = \infty$, and $\lim_{y \rightarrow \infty} \pi_H(y) = 0$ can be leveraged to find an identical expression for the average time in the high patch:

$$\bar{T}_H(y) = \frac{\mathcal{T}_H(y)}{\pi_H(y)} = \frac{y - \theta}{(\lambda_H - \lambda_L) - \lambda_L \log \frac{\lambda_H}{\lambda_L}}.$$

For the Bayesian model, the prior and initial condition of a patch is $y(0) = \log \frac{p_H}{1-p_H}$, so the relevant quantities for our problem are $\pi_L \equiv 1$; $\pi_H = \frac{1-p_H}{p_H} e^\theta$; and

$$\bar{T}_L = \bar{T}_H = \frac{\log \frac{p_H}{1-p_H} - \theta}{(\lambda_H - \lambda_L) - \lambda_L \log \frac{\lambda_H}{\lambda_L}}.$$

Substituting these expressions into Eq. (5), we obtain an explicit formula for \bar{T}_{arrive} as it depends on model parameters (Fig. 3A,B). Minimizing \bar{T}_{arrive} to find θ^{opt} then requires solving:

$$\operatorname{argmin}_{\theta \in (-\infty, \log \frac{p_H}{1-p_H})} (1 - p_H) \left[\frac{1 + e^\theta}{p_H - (1 - p_H)e^\theta} \right]$$

$$\times \left[\frac{\log \frac{p_H}{1-p_H} - \theta}{(\lambda_H - \lambda_L) - \lambda_L \log \frac{\lambda_H}{\lambda_L}} + \tau \right]$$

whose solutions θ^{opt} are identified with critical points obeying $\theta < \log \frac{p_H}{1-p_H}$ and

$$e^\theta \left[\log \frac{p_H}{1-p_H} - \theta + A\tau \right] = (1 + e^\theta)(p_H - (1 - p_H)e^\theta), \quad (11)$$

where $A = (\lambda_H - \lambda_L) - \lambda_L \log \frac{\lambda_H}{\lambda_L}$, which we can solve for θ^{opt} numerically (solid lines in Fig. 3C,D). An explicit approximation of θ^{opt} is obtained by dropping the $e^{2\theta}$ term (expecting θ negative) and solving to find

$$\theta^{\text{opt}} \approx W_{-1} \left[-(1 - p_H) \left(\frac{\lambda_H}{\lambda_L} \right)^{\lambda_L \tau} e^{-(\lambda_H - \lambda_L)\tau} e^{2p_H - 1} \right] + A\tau + \log \frac{p_H}{1-p_H} + 1 - 2p_H, \quad (12)$$

where $W_{-1}(z)$ is the $(-1)^{\text{th}}$ branch of the Lambert W function (inverse of $z = We^W$). This approximation compares well with numerical solutions to Eq. (11) (dotted lines in Fig. 3C,D), and can be further simplified using the approximation $W_{-1}(z) \approx \log(-z) - \log(-\log(-z))$ yielding

$$\theta^{\text{opt}} \approx \log(1 - p_H) - \log[-\log(1 - p_H) + (\lambda_H - \lambda_L)\tau]$$

$$-\lambda_L \tau \log \left[\frac{\lambda_H}{\lambda_L} + 1 - 2p_H \right], \quad (13)$$

from which we can observe how the optimal threshold scales in limits of environmental parameters (dashed lines in Fig. 3C,D).

Examining Eq. (12), we see that when high yield patches are rare ($p_H \rightarrow 0$), the optimal threshold diverges as $\theta^{\text{opt}} \rightarrow -\infty$ with the scaling $\log p_H$, since the initial condition becomes more negative. When low yield patches are rare and $p_H \rightarrow 1$, $\theta^{\text{opt}} \rightarrow -\infty$ since the first selected patch will almost always be high yielding. Therefore, if high yielding patches are very rare or common, the threshold should be moved far from zero to require high certainty when abandoning a patch. Between, $\theta^{\text{opt}} > -\infty$ varies nonmonotonically in p_H . For large $\lambda_H/\lambda_L \gg 1$, $\theta^{\text{opt}} \sim -\log(\lambda_H)$, since patches are easier to distinguish as λ_H increases, so again high certainty can be required to depart a patch. In the limit $\lambda_H \rightarrow \lambda_L$, $\theta^{\text{opt}} \approx \log(1-p_H) - \log[-\log(1-p_H) + 1 - 2p_H]$. Note, when both increasing discriminability (λ_H/λ_L) and the high yield patch fraction p_H , the minimal mean time $T_{\text{arrive}}^{\text{opt}}$ needed to arrive and remain in a high yield patch decreases (Fig. 3E,F).

Thus, the problem of patch leaving can be reduced to a threshold crossing process in the case of a binary non-depleting environment. Indeed, we find the model is tractable enough to identify explicit scaling relations between model parameters and decision strategies. Namely, the time to arrive and remain in the high yielding patch increases as high patches become less discriminable (λ_H/λ_L close to one) and more rare (small p_H). Patch departures require more certainty when high patches are more discriminable and/or are very rare or common. Next, we extend our study of the high patch arrival problem to non-depleting environments with more than two patch types. Again, we find this can be accomplished by setting a threshold on the belief about whether or not the agent is in the highest yielding patch or not.

B. Three patch types

In ternary environments, there are three patch arrival rates $\lambda_1 > \lambda_2 > \lambda_3 \geq 0$ with $p_0(\lambda) = \sum_{j=1}^3 p_j \delta(\lambda - \lambda_j)$ and $\sum_{j=1}^3 p_j = 1$ (Fig. 4A). Representing log-likelihoods using Eq. (2) and defining LLRs $y_1 = \log \frac{p(\lambda_1|x(t))}{p(\lambda_2|x(t))}$ and $y_2 = \log \frac{p(\lambda_1|x(t))}{p(\lambda_3|x(t))}$ for the beliefs within a patch given the food encounter time series $x(t) = \sum_{j=1}^{K(t)} \delta(t-t_j)$, we have the planar system

$$y_1' = \log \frac{\lambda_1}{\lambda_2} \sum_{j=1}^{K(t)} \delta(t-t_j) - (\lambda_1 - \lambda_2), \quad (14a)$$

$$y_2' = \log \frac{\lambda_1}{\lambda_3} \sum_{j=1}^{K(t)} \delta(t-t_j) - (\lambda_1 - \lambda_3), \quad (14b)$$

where $y_1(0) = \log \frac{p_1}{p_2}$ and $y_2(0) = \log \frac{p_1}{p_3}$. Note that we can recover any of the three likelihoods from y_1 and y_2 by the following mappings

$$p(\lambda_{j+1}|x(t)) = \frac{e^{-y_j}}{1 + e^{-y_1} + e^{-y_2}},$$

where $y_0 = 0$ for $j = 0$. An argument similar to the binary case demonstrates that the best patch leaving policies in ternary environments cause the forager to find and remain in the highest yielding patch (λ_1). This is accomplished by thresholding the probability of being in the high yielding patch, so when $p(\lambda_1|x(t)) = \phi \in (0, p_1)$, the forager exits the patch. In (y_1, y_2) space, this threshold is a parameterized curve

$$y_2^\phi = -\log \left[\frac{1-\phi}{\phi} - e^{-y_1^\phi} \right], \quad y_1^\phi \in \left(-\log \frac{1-\phi}{\phi}, \infty \right), \quad (15)$$

bounded by $\lim_{y_j^\phi \rightarrow \infty} y_k^\phi = -\log \frac{1-\phi}{\phi} \equiv \theta$, suggesting we approximate the boundary in Eq. (15) with $y_1, y_2 \geq \theta$. This leads to the forager departing the patch given sufficient evidence they are not in the highest yielding patch (Fig. 4B).

Given such a strategy, the mean time \bar{T}_{arrive} to arrive and remain in the high yielding patch is computed from the escape probability $\pi_1(\theta)$ from the high yielding patch and the mean times to visit each patch $\bar{T}_j(\theta)$ ($j = 1, 2, 3$) when escaping. As before, the forager will always escape the lower yielding patches (2 and 3). The mean arrival time is then given by

$$\bar{T}_{\text{arrive}}(\theta) = \frac{\pi_1(\theta)}{1 - \pi_1(\theta)} (\bar{T}_1(\theta) + \tau) + \frac{1 - p_1}{p_1} \frac{p_2(\bar{T}_2(\theta) + \tau) + p_3(\bar{T}_3(\theta) + \tau)}{1 - \pi_1(\theta)}. \quad (16)$$

While we could compute the needed MFPT statistics for Eq. (16) by solving a corresponding backward Kolmogorov equation, the resulting two-dimensional delayed PDE proves more computationally expensive to solve than Monte Carlo sampling Eq. (14). We thus save a more detailed asymptotic analysis of higher order PDEs associated with multi-patch problems for future work, and report results from sampling here (Fig. 4C-F). The mean high patch arrival time \bar{T}_{arrive} depends strongly on the high patch food arrival rate λ_1 , decreasing considerably as the patch becomes more discriminable (Fig. 4C). On the other hand, \bar{T}_{arrive} depends weakly on the worst patch's food arrival rate λ_3 (Fig. 4D). The discriminability of λ_2 and λ_3 does not strongly affect the observer's ability to determine whether a patch is high yielding (λ_1) or one of the two lower yielding patches (λ_2 or λ_3) as long as λ_1 and λ_2 are sufficiently discriminable. In a related way, \bar{T}_{arrive} is much more strongly affected by changes in the fraction of the high yielding patch (p_1 : Fig. 4E) than changes in the balance of the middle (λ_2) and low (λ_3) patches (Fig. 4F).

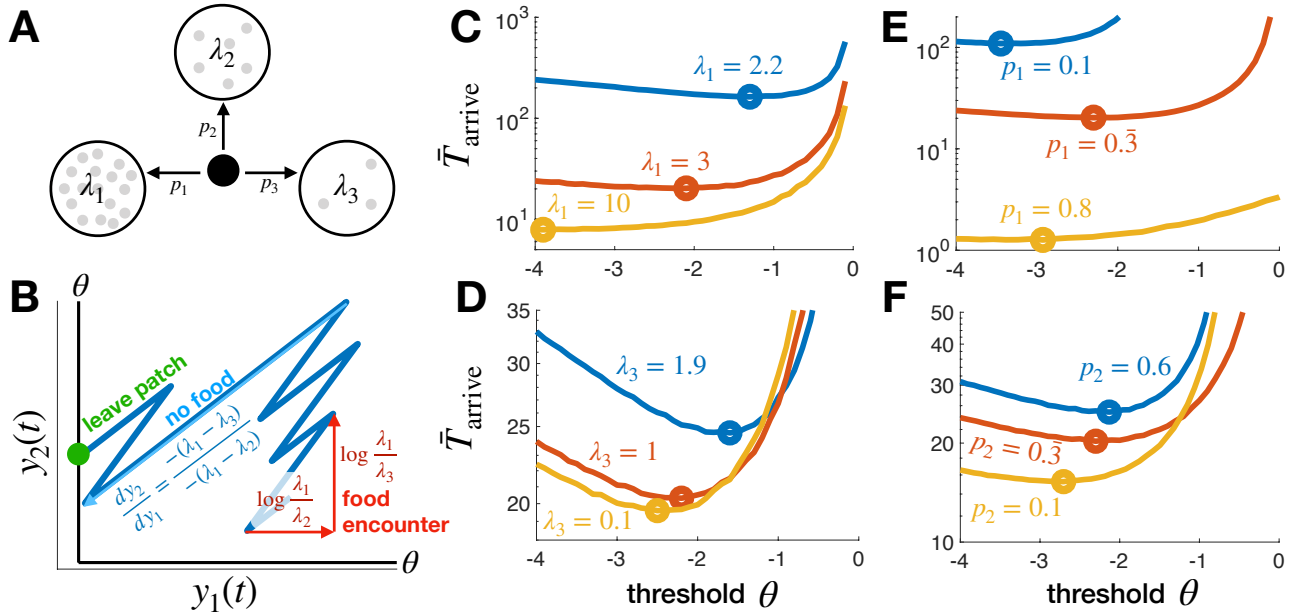


FIG. 4. Ternary (3 patch) non-depleting environments. **A.** Distribution of three possible patches. **B.** Patch beliefs $y_1(t) = \log \frac{p(\lambda_1|x(t))}{p(\lambda_2|x(t))}$ and $y_2(t) = \log \frac{p(\lambda_1|x(t))}{p(\lambda_3|x(t))}$ increase with food encounters and decrease between until either reaches the departure threshold θ . **C.** Mean time \bar{T}_{arrive} to arrive and remain in highest yielding patch λ_1 (Eq. (16)) decreases with patch discriminability λ_1 as does the optimal departure threshold θ^{opt} (circles). Other parameters are $\lambda_2 = 2$, $\lambda_3 = 1$, $\tau = 5$, and $p_1 = p_2 = p_3 = 1/3$. **D.** \bar{T}_{arrive} increases with λ_3 as does θ^{opt} , since the worst patches become less easy to distinguish from the best ($\lambda_1 = 3$). Other parameters as before. **E.** \bar{T}_{arrive} decreases with the prevalence of the best patch p_1 (while $p_2 = p_3 = (1 - p_1)/2$), but θ^{opt} varies non-monotonically. **F.** $\bar{T}_{\text{arrive}}^{\text{opt}}$ increases with p_2 (while $p_1 = 0.333$ and $p_3 = 1 - p_1 - p_2$) as does θ^{opt} . All curves for \bar{T}_{arrive} computed from 10^6 Monte Carlo simulations.

Importantly, the optimal threshold θ^{opt} decreases as patches become more easily discriminable (Fig. 4C,D), but can vary non-monotonically with the fraction p_1 of high patches as in the binary environment (Fig. 3D). Increases in θ^{opt} at lower values of p_1 arise from the fact that departure from high valued patches will be interspersed with long sequences of visits to lower yielding patches. Thus, there is a higher premium on distinguishing the high yielding patch when actually in one, similar to how subjects raise their decision thresholds when inter-trial intervals are lengthened in two alternative forced choice paradigms [55]. At higher values of p_1 , one is more likely to land in a high yielding patch, so again one can afford to require more certainty to depart.

We now extend our analysis of departure strategies to environments with $N > 3$ patch types. As before, we expect that patch discrimination will mainly be impacted by the spacing between the highest and second highest food arrival rates λ_1 and λ_2 , as we now show in detail.

C. More patch types

Consider environments with N patch types having arrival rates $\lambda_1 > \lambda_2 > \dots > \lambda_N \geq 0$ with $p_0(\lambda) = \sum_{j=1}^N p_j \delta(\lambda - \lambda_j)$ and $\sum_{j=1}^N p_j = 1$ (Fig. 5A). In this case, defining the LLRs $y_j = \log \frac{p(\lambda_1|x(t))}{p(\lambda_{j+1}|x(t))}$ for $j = 1, \dots, N-1$,

yields the $N - 1$ -dimensional system

$$y'_j = \log \frac{\lambda_1}{\lambda_{j+1}} \sum_{j=1}^{\infty} \delta(t - t_j) - (\lambda_1 - \lambda_{j+1}), \quad (17)$$

where $y_j(0) = \log \frac{p_1}{p_{j+1}}$, and any likelihood can be recovered as

$$p(\lambda_{j+1}|x(t)) = \frac{e^{-y_j}}{1 + \sum_{k=1}^{N-1} e^{-y_k}},$$

where $y_0 = 0$ for $j = 0$.

A full analysis of optimal strategies across a broad range of parameters is prohibitive, so we focus on the impact of varying a single departure threshold θ on the time to arrive and remain in a high yielding patch. Noting the weak dependence of arrival times on parameters for the low patch (p_3 and λ_3) in the three patch case, we consider simplified strategies in which the observer only tracks the first L LLRs y_1, y_2, \dots, y_L and compares these with the threshold θ to decide when to leave the patch (Fig. 5B). Thus, we compute the mean time to arrive and remain in the high patch:

$$\bar{T}_{\text{arrive}}(\theta) = \frac{\pi_1(\theta, L)}{1 - \pi_1(\theta, L)} (\bar{T}_1(\theta, L) + \tau) + \frac{1 - p_1}{p_1} \frac{\sum_{j=2}^N p_j (\bar{T}_j(\theta, L) + \tau)}{1 - \pi_1(\theta)}, \quad (18)$$

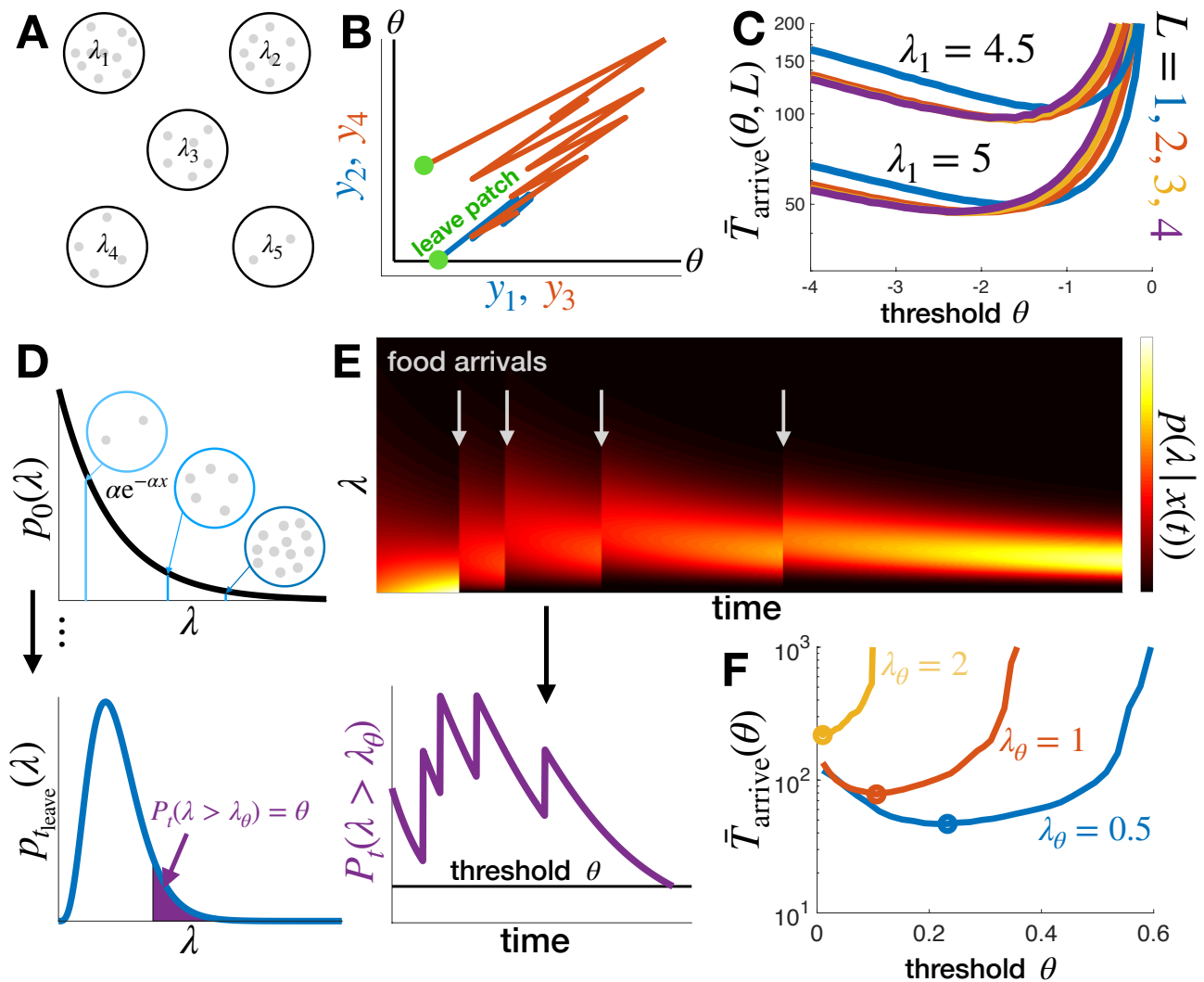


FIG. 5. Finding high yielding patches among many patch types. **A.** Environment with 5 patch types. **B.** Patch beliefs $y_j(t) = \log \frac{p(\lambda_1|x(t))}{p(\lambda_{j+1}|x(t))}$ evolve on an $N - 1$ -dimensional space until one y_j ($j = 1, \dots, L \leq N - 1$) reaches θ . **C.** The mean time \bar{T}_{arrive} to arrive and remain in highest yielding patch λ_1 as computed in Eq. (18) slightly decreases as more LLRs (L increases) are used ($p_j = 1/N$, $\lambda_j = 6 - j$, $j = 1, 2, 3, 4$, $N = 5$). **D.** Continuum of patch types given by an exponential prior $\lambda \sim \exp[\alpha]$. The forager departs a patch when the probability that the arrival rate is above threshold ($\lambda > \lambda_\theta$) falls to a threshold value $P_t(\lambda > \lambda_\theta) = \theta$. **E.** The posterior $p(\lambda|x(t))$ of the patch arrival rate is shifted up by food arrives, while it decreases in the time between. The mass $P_t(\lambda > \lambda_\theta)$ follows the same pattern. **F.** There is an optimal θ that minimizes the time to arrive and remain in a high yielding patch ($\lambda > \lambda_\theta$), and the optimal time $\bar{T}_{\text{arrive}}^{\text{opt}}$ (circles) increases as the acceptable arrival rate is increased. $\alpha = 1$ is used here. In **C** and **F**, 10^7 Monte Carlo simulations are used to compute the curves \bar{T}_{arrive} .

where patch departure strategy depends on the number L of LLRs thresholded and the threshold θ used. In a five patch type example, performance depends weakly on $L > 2$, and it is sufficient to simply track the LLRs between the first three patches (Fig. 5C).

D. Continuum limit

Next, consider the limit of many patches, in which patch arrival rate λ is drawn from a continuous distribution $p_0(\lambda)$ defined on $\lambda \in [0, \infty)$ (Fig. 5D). This gen-

eral case is described in Eq. (2). Rather than resorting to LLRs, we retain the posterior $p(\lambda|x(t))$, and compute the mass of a thresholded portion of this pdf. Given a reference arrival rate λ_θ , we track $P(\lambda > \lambda_\theta|x(t)) = \int_{\lambda_\theta}^{\infty} p(\lambda|x(t))d\lambda$. This is analogous to a forager who seeks patches with yield rates λ_θ or above, but deems lower yield rates to be insufficient. For any continuous pdf $p_0(\lambda)$, the maximum λ will never be sampled, so arriving and remaining in the true maximum yielding patch is infeasible.

For the case of an exponential prior $p_0(\lambda) = \alpha e^{-\alpha\lambda}$,

given $K(t)$ food arrivals, we can compute

$$\begin{aligned} P(\lambda > \lambda_\theta | x(t)) &= \frac{1}{p(x(t))} \int_{\lambda_\theta}^{\infty} \alpha e^{-\alpha \lambda} \lambda^{K(t)} e^{-\lambda t} d\lambda \\ &= \frac{\alpha}{(\alpha + t)^{K(t)+1} p(x(t))} \int_{(\alpha+t)\lambda_\theta}^{\infty} \nu^{K(t)} e^{-\nu} d\nu \\ &= \frac{\alpha}{(\alpha + t)^{K(t)+1} p(x(t))} \Gamma(K(t), (\alpha + t)\lambda_\theta), \end{aligned}$$

where $\Gamma(s, x)$ is the incomplete gamma function, and in a similar way

$$P(\lambda < \lambda_\theta | x(t)) = \frac{\alpha}{(\alpha + t)^{K(t)+1} p(t_{1:K(t)})} \gamma(K(t), (\alpha + t)\lambda_\theta),$$

where $\gamma(s, x)$ is the lower incomplete gamma function. As such, we can compute the LLR

$$\rho(t) = \log \frac{P(\lambda > \lambda_\theta | x(t))}{P(\lambda < \lambda_\theta | x(t))} = \log \frac{\Gamma(K(t), (\alpha + t)\lambda_\theta)}{\gamma(K(t), (\alpha + t)\lambda_\theta)}, \quad (19)$$

and state the forager departs the patch when $\rho(t) \leq \hat{\theta}$ or when $P(\lambda > \lambda_\theta | x(t)) \leq \theta := 1/(1 + e^{-\hat{\theta}})$ (Fig. 5D,E). Note, to allow evidence accumulation, we require that $\theta < \alpha \int_{\lambda_\theta}^{\infty} e^{-\alpha \lambda} d\lambda = e^{-\alpha \lambda_\theta} \equiv \phi$, which represents the fraction of patches where $\lambda \geq \lambda_\theta$.

As with other cases, we can compute the mean time to find and remain in a patch of high enough quality. In fact, we can marginalize over all patch types of each class (high and low) to compute the probability of escaping a high patch across all high patches,

$$\pi_H(\theta; \lambda_\theta) = \int_{\lambda_\theta}^{\infty} p_0(\lambda) \pi(\theta; \lambda) d\lambda,$$

and the mean time per visit to a high and low patch types,

$$\bar{T}_H(\theta; \lambda_\theta) = \int_{\lambda_\theta}^{\infty} p_0(\lambda) \bar{T}(\theta; \lambda) d\lambda,$$

$$\bar{T}_L(\theta; \lambda_\theta) = \int_0^{\lambda_\theta} p_0(\lambda) \bar{T}(\theta; \lambda) d\lambda,$$

when departing across all patches of each type. We compute the integrals above via Monte Carlo sampling. Defining $p_H = \int_{\lambda_\theta}^{\infty} p_0(\lambda) d\lambda$, the mean arrival time is then given by the two patch formula, Eq. (5), as we have partitioned the environment into two patch categories.

As the observer places a higher threshold λ_θ on the quality of an acceptably high yielding patch, the optimal time to arrive in such a patch increases (Fig. 5F). Moreover, the optimal threshold θ decreases, as more time must be spent in patches to properly discriminate a high yielding patch, as these become rarer in the prior as λ_θ increases. In connection with the binary environment, increasing λ_θ corresponds to making high patches more discriminable but also more rare.

Our analysis of patch departure strategies in the context of non-depleting patches reveals a number of consistent trends across patch type counts. First, the optimal time to arrive and remain in the highest yielding patch decreases as the high patch discriminability increases and as high patches become more common. Second, in environments with more than two patch types, the most important parameters in determining the time to find the highest yielding patch are those related to the highest and second highest yielding patch types. To most efficiently find the high patch, it is sufficient to compute LLRs corresponding to the first two or three patch types. In this regard, we expect that reasonably effective strategies for foraging environments with a continuum of patch types could be generated using particle filters that only compute likelihoods over a finite sample of possible patch types [56]. Strategies for non-depleting environments involve minimizing the time to find a high yielding patch. In depleting environments, the forager eventually must leave any patch it arrives in, as food sources are exhausted. We study such strategies in the next section.

IV. DEPLETION- VS. UNCERTAINTY-DRIVEN DECISIONS IN DEPLETING ENVIRONMENTS

Foraging animals deplete resources as they use them [57, 58]. This is accounted for in the most general form of the foraging agent's belief in Eq. (1), and thus far we have analyzed idealized versions of the model that neglect this. To understand the impact of depletion on the agent's belief and strategy, we will focus here on simple cases: (a) homogeneous environments in which all patches are the same and the forager knows the initial arrival rate λ_0 a priori; (b) binary environments in which the forager knows there are two types of patches, and may or may not have to infer which of the two they are currently foraging in while they deplete it; and (c) environments with a few patches in which there is a memory of depletion from previous patch visits.

A. Homogeneous environments

We start by considering an agent foraging in an environment in which all patches begin with the same density of food, so the arrival rate in each is initially λ_0 . As they forage, the arrival rate $\lambda(t) = \lambda_0 - K(t)\rho$ decreases in equal increments with each food encounter (Fig. 6A). We assume the agent perfectly knows this arrival rate as well as the amount of time it has been foraging, and will use this information to determine a strategy for departing the patch. In keeping with the simplicity of the space of possible strategies discussed before, we will assume the forager departs the patch when $\lambda(t)$ (or analogously the density of food in the patch) falls below some threshold λ_θ .

This problem can be framed by simply computing the

mean food intake rate for a given strategy parameterized by λ_θ . Assuming λ_θ is an integer multiple of ρ , we can compute directly the number of chunks consumed before departure $m_\theta \equiv K(T(\lambda_\theta)) := (\lambda_0 - \lambda_\theta)/\rho$, and the mean time to depart $\bar{T}(\lambda_\theta)$. Linearity of expectations allows us to compute the mean departure time as the sum of mean exponential waiting times between food encounters

$$\begin{aligned}\bar{T}(\lambda_\theta) &= \sum_{j=1}^{m_\theta} \frac{1}{\lambda_0 - (j-1)\rho} = \frac{1}{\rho} \sum_{j=1}^{m_\theta} \frac{1}{m_0 - (j-1)} \\ &= \frac{H_{m_0} - H_{m_0 - m_\theta}}{\rho},\end{aligned}$$

where H_n is the n th harmonic number. Thus, we compute the long term reward rate given λ_θ as

$$R^{\lambda_\theta} = \frac{m_0\rho - \lambda_\theta}{H_{m_0} - H_{\lambda_\theta/\rho} + \rho\tau}. \quad (20)$$

There is an interior optimum m_θ that maximizes long term food consumption rate (Fig. 6B,C). For low initial rate λ_0 and decrement ρ , the best strategy is to remain in a patch until all food is consumed, but as λ_0 and ρ are increased, the optima occur roughly where $R^{\lambda_\theta} \equiv \lambda_\theta$ as in the marginal value theorem (MVT).

To analyze this observation further, we can approximate optima in the limit of plentiful patches ($m_0 \gg 1$), such that H_n can be estimated by the large n expansion

$$H_n = \log n + \gamma + \mathcal{O}(1/n), \quad (21)$$

where $\gamma = 0.57721566\dots$ (the Euler–Mascheroni constant). Using this, a first approximation yields

$$R^{\lambda_\theta} \approx \frac{m_0\rho - \lambda_\theta}{\log(\rho m_0) - \log \lambda_\theta + \rho\tau}. \quad (22)$$

In this limit, we can explicitly show the MVT is satisfied, $R^{\lambda_\theta} \equiv \lambda_\theta$, and the observer departs the patch when the the arrival rate of food in the patch matches their average environment-wide food arrival rate. To show this, note that R^{λ_θ} in Eq. (22) has derivative

$$\frac{dR^{\lambda_\theta}}{d\lambda_\theta} = \frac{-(\log(\rho m_0) - \log \lambda_\theta + \tau/\rho) + (m_0\rho - \lambda_\theta)/\lambda_\theta}{(\log(\rho m_0) - \log \lambda_\theta + \rho\tau)^2},$$

which is increasing at $\lambda_\theta = \rho$:

$$\left. \frac{dR^{\lambda_\theta}}{d\lambda_\theta} \right|_{\lambda_\theta=\rho} = \frac{(m_0 - 1) - (\log(m_0) + \tau/\rho)}{(\log m_0 + \rho\tau)^2} > 0,$$

since $m_0 \gg 1$, and decreasing at $\lambda_\theta = \rho m_0 = \lambda_0$:

$$\left. \frac{dR^{\lambda_\theta}}{d\lambda_\theta} \right|_{\lambda_\theta=\lambda_0} = -\frac{1}{\rho\tau} < 0.$$

In between lies a critical point satisfying

$$\frac{m_0\rho - \lambda_\theta}{\lambda_\theta} = \log \frac{\rho m_0}{\lambda_\theta} + \rho\tau, \quad (23)$$

which is precisely the equation recovered when setting $R^{\lambda_\theta} \equiv \lambda_\theta$ in Eq. (22), implying the critical point of R^{λ_θ} occurs where $R^{\lambda_\theta} \equiv \lambda_\theta$ as in the MVT.

Alternatively, for small m_0 (few chunks of food per patch), computing the harmonic numbers in Eq. (20) is straightforward, so we can directly compare the mean rate R^{λ_θ} for all possible values of λ_θ . For instance, when $m_0 = 2$, we can choose from $\lambda_\theta \in \{0, \rho\}$ as the departure threshold, corresponding to maximizing

$$R^{\lambda_\theta} \in \left\{ \frac{4\rho}{3 + 2\rho\tau}, \frac{2\rho}{1 + 2\rho\tau} \right\} = \{R^{\lambda_\theta=0}, R^{\lambda_\theta=\rho}\}.$$

With this, the low threshold should be chosen when $\rho > 1/(2\tau)$, or when energy gained from food chunks is larger than half the transit rate between patches. At $\rho = 1/(2\tau)$, the forager should leave after a single chunk is consumed, yielding $R^{\text{opt}} = \rho = \lambda_\theta^{\text{opt}}$ (as in the MVT). Similar results can be computed in the cases $m_0 = 3, 4, 5, \dots$ using explicit computation of the harmonic numbers H_n . We expect these results to generalize to arbitrary m_0 as evidenced by the small m_0 case and our large m_0 asymptotics.

Overall this suggests that for short transit times τ or small food chunk sizes ρ , and ideal observer will depart patches more quickly than when transit times are long τ or food chunks have a higher quality ρ . This is consistent with classic models of continuous consumption as in the MVT, but here we have provided a detailed analysis of the case of discretized and random food encounters. The primary reason for a mismatch with the MVT arises due to the discreteness of the food availability and arrival rate. We show in binary environments, when uncertainty plays a role in patch leaving decisions, there are nontrivial deviations from the MVT due to the forager not precisely knowing the food arrival rate of their current patch.

B. Binary environments

Next, we consider binary depleting environments in which the observer must both infer the type of patch (high or low) and track its depletion. We focus on strategies in which the observer departs the patch when the mean estimate of the arrival rate $\bar{\lambda}(t)$ falls below a threshold λ_θ . To understand this most general case, we begin by discussing two simplified cases: (a) the case in which the patch type is known on arrival, and (b) the case in which the low yielding patch is empty ($m_0^L = 0$).

Depletion-dominated regime. One way of contextualizing the case of known patch types is that of depletion-dominated decisions. These types of decisions occur if the observer stays in a patch till they have determined the patch type to high certainty. To idealize this situation, assume the observer arrives in a patch and immediately knows the patch type λ_j ($j \in \{H, L\}$) in which they reside. Moreover, assume they leverage a strategy in which they depart the patch as soon as the arrival rate falls below some level λ_θ^j . In this case, there are two thresholds

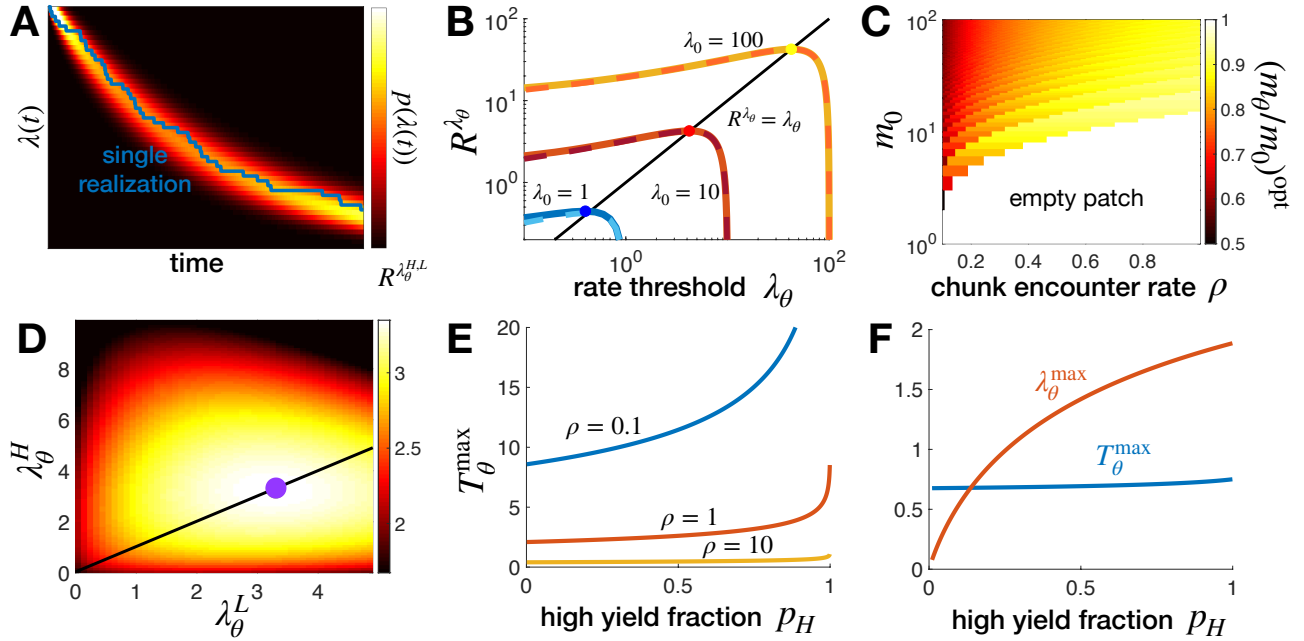


FIG. 6. Patch leaving strategies in a depleting environment. **A.** In homogeneous environments where all patches have initial arrival rate λ_0 , the forager maintains full knowledge of the current arrival rate by counting food encounters. **B.** Rate of food consumption as a function of strategy R^{λ_θ} , where the forager departs once arrival rate falls to or below λ_θ . Solid lines are exact solution of Eq. (20), and dashed lines are the large m_0 approximation Eq. (22). Note optimal $\lambda_\theta \approx R^{\lambda_\theta}$. **C.** Optimal m_θ/m_0 as a function of ρ and m_0 . **D.** Rate of food consumption $R^{\lambda_{\theta^{H,L}}}$ in binary depleting environments in which the observer knows patch type (λ_H or λ_L) upon arrival. Optimal strategy (purple dot) takes $\lambda_\theta^H = \lambda_\theta^L \approx R^{\lambda_{\theta^{H,L}}} \approx 3.3$. $p_H = 0.5$, $m_0^H = 100$, $m_0^L = 50$, and $\rho = 0.1$. **E.** Optimal waiting time T_θ^{\max} in environments where high (low) patch has one (zero) chunk(s) of food. **F.** Optimal waiting time T_θ^{\max} and departure threshold λ_θ^{\max} in a binary environment where the low patch has zero chunks of food and the high patch initially has $m_0^H = 50$. $\rho = 0.1$. Transit time $\tau = 5$ throughout.

to tune for each of the patch types. Following our calculations from the homogeneous case, the time to depart a patch of type j is

$$\bar{T}_j(\lambda_\theta) = \frac{H_{m_0^j} - H_{m_0^j - m_\theta^j}}{\rho},$$

where m_0^j is the initial food count in patches of type j and m_θ^j is the threshold level of food to consume before departing. As such, the long term reward rate is

$$R^{\lambda_{\theta^{H,L}}} = \frac{p_H(m_0^H \rho - \lambda_\theta^H) + p_L(m_0^L \rho - \lambda_\theta^L)}{p_H(H_{m_0^H} - H_{\lambda_\theta^H/\rho}) + p_L(H_{m_0^L} - H_{\lambda_\theta^L/\rho}) + \rho\tau}, \quad (24)$$

maximized by setting $\lambda_\theta^H = \lambda_\theta^L \approx R^{\lambda_{\theta^{H,L}}}$ (Fig. 6D). We show now that this follows the MVT.

In large m_0^j limit, the harmonic numbers are approximated by Eq. (21), and Eq. (24) becomes

$$R^{\lambda_{\theta^{H,L}}} \approx \frac{p_H(m_0^H \rho - \lambda_\theta^H) + p_L(m_0^L \rho - \lambda_\theta^L)}{p_H \tilde{T}_H(\lambda_\theta^H) + p_L \tilde{T}_L(\lambda_\theta^L) + \rho\tau}, \quad (25)$$

where $\tilde{T}_j(\lambda_\theta^j) = \log(\rho m_0^j) - \log \lambda_\theta^j$ and the critical point equations for each partial derivative $R_{\lambda_\theta^j}^{\lambda_{\theta^{H,L}}} = 0$ imply

$$p_H(m_0^H - \lambda_\theta^H) + p_L(m_0^L \rho - \lambda_\theta^L) = \lambda_\theta^j [p_H(\log(\rho m_0^H)$$

$$- \log \lambda_\theta^H) + p_L(\log(\rho m_0^L) - \log \lambda_\theta^L) + \rho\tau], \quad (26)$$

which can be rewritten as $R^{\lambda_{\theta^{H,L}}} = \lambda_\theta^j$ ($j = H, L$), implying the observer should depart a patch when the arrival rate equals the mean rate of food arrival for the environment (MVT). As in the homogeneous depleting environment, obtaining the prediction of the MVT relies on the observer knowing the current food arrival rate. However, as we will show, the optimal strategy deviates from the MVT when there is uncertainty concerning the present food arrival rate in the patch.

Empty low patch. Before addressing the general case of a binary environment with initially unknown patch types, we study a simple case in which the forager does not know the patch type they are in ahead of time but where the statistics of patch departures are still explicitly calculable: when the low yielding patch is empty.

The simplest case is that in which some patches are empty ($m_0^L = 0$) and others have a single chunk of food ($m_0^H = 1$), as introduced by [36] but not analyzed in detail. Clearly, when a forager encounters a chunk of food they should leave the patch, since there is no food left thereafter. How long should they wait until departing a patch if food has yet to be encountered? This amounts to optimizing the waiting time T_θ till departure (or equivalently setting a threshold mean estimated arrival rate λ_θ at which they depart). Each patch visit thus falls into

one of three categories: (a) visits to empty (low) patches lasting time T_θ ; (b) visits to high patches resulting in no food lasting time T_θ ; and (c) visits to high patches resulting in one food chunk lasting time $t < T_\theta$.

To determine an optimal strategy, we need only compute statistics for case (c). Since the food arrival rate in a high patch is ρ , the probability of encountering food when waiting a time T_θ is $1 - \pi_\theta = 1 - e^{-\rho T_\theta}$, the mean time to wait is

$$\bar{T}_H^\theta = \frac{\rho \int_0^{T_\theta} t e^{-\rho t} dt}{1 - e^{-\rho T_\theta}} = \frac{1 - (\rho T_\theta + 1)e^{-\rho T_\theta}}{\rho(1 - e^{-\rho T_\theta})},$$

and the rate of consumption of a given strategy is

$$R^\theta = \frac{p_H(1 - \pi_\theta)}{p_H(\pi_\theta T_\theta + (1 - \pi_\theta)\bar{T}_H^\theta) + p_L T_\theta + \tau},$$

which we can maximize by finding the critical point of R^θ in T_θ . As p_H is increased and as ρ is decreased, the optimal $T_\theta = T_\theta^{\max}$ increases (Fig. 6E). When high yielding patches are more frequent (higher p_H), one should stay in a patch longer, since they are more likely to encounter food. When the rate of chunk discovery ρ is smaller, one should expect to wait longer to encounter food in a high patch, so the wait time should be increased.

Note, the observer does not leave the initially empty patches immediately, so their strategy (even if optimally tuned) does not agree with the MVT due to the observer's uncertainty about each patch's yield rate.

We also analyze the case of an empty low patch and an arbitrary number of food chunks m_0 in the high patch. Here, the observer should apply a hybrid strategy: Wait a finite time T_θ to depart if no food is encountered, but if food is encountered before $t = T_\theta$, consume food until the arrival rate drops to $\lambda_\theta = (m_0 - m_\theta)\rho$. This strategy distinguishes two types of patch leaving decisions arising in uncertain and depleting environments. Decisions may be uncertainty-dominated, where the observer may depart a patch early, believing they are in a low-yielding patch. Alternatively, decisions may be depletion-dominated, after the observer is nearly certain of they patch type they are in and has depleted it.

To optimally parameterize the departure strategy, we again compute the probability of departing the high patch early $\pi_\theta = e^{-m_0\rho T_\theta}$ (at time T_θ). Assuming the observer remains in the patch, it takes a time

$$\bar{T}_H^\theta = \frac{1 - (m_0\rho T_\theta + 1)e^{-m_0\rho T_\theta}}{m_0\rho(1 - e^{-m_0\rho T_\theta})}$$

to consume the first chunk of food and then depart immediately ($T(\lambda_\theta) = 0$) if $\lambda_\theta = \lambda_0 - \rho$ (as in the previous case). Otherwise, it will take an additional time

$$\begin{aligned} \bar{T}(\lambda_\theta) &= \sum_{j=2}^{m_\theta} \frac{1}{\lambda_0 - (j-1)\rho} = \frac{1}{\rho} \sum_{j=1}^{m_\theta-1} \frac{1}{m_0 - j} \\ &= \frac{H_{m_0-1} - H_{m_0-m_\theta}}{\rho}. \end{aligned}$$

At each patch then, the observer either obtains no food and departs after time T_θ or obtains $m_\theta = m_0 - \lambda_\theta/\rho$ chunks and departs on average after time $\bar{T}_H^\theta + T(\lambda_\theta)$. The probability of the latter is $p_H(1 - \pi_\theta)$, and assuming $m_0 \gg 1$, we can approximate $H_{m_0-1} - H_{m_0-m_\theta} \approx \log(m_0 - 1) - \log(m_0 - m_\theta)$, so the food consumption rate is approximated by the a continuous function

$$R(T_\theta, \lambda_\theta) \approx \frac{p_H(1 - \pi_\theta)(m_0\rho - \lambda_\theta)}{p^\theta(\bar{T}(\lambda_\theta) + \bar{T}_H^\theta) + (1 - p^\theta)\rho T_\theta + \rho\tau},$$

where $p^\theta = p_H(1 - \pi_\theta)$ and $\bar{T}(\lambda_\theta) = \log(\rho m_0 - \rho) - \log \lambda_\theta$. Reward rate is maximized using a waiting time T_θ that is fairly insensitive to p_H , whereas the threshold λ_θ is sensitive to p_H (Fig. 6F). When high yielding patches become more common, the forager need not deplete them as much, since the next patch they visit is likely to be high yielding.

Many food chunks. Lastly, we examine general binary depleting environments in which $m_0^H > m_0^L$ are arbitrary integers, so the belief $y(t) = \log \frac{P(\lambda_H - K(t)\rho|x(t))}{P(\lambda_L - K(t)\rho|x(t))}$ evolves according to Eq. (4). In line with our previous analysis, the observer estimates the current arrival rate of the patch from this belief,

$$\tilde{\lambda}(t) = \frac{\lambda_H + e^{-y}\lambda_L}{1 + e^{-y}} - \rho K(t), \quad (27)$$

and departs the patch when $\tilde{\lambda}(t) \leq \lambda_\theta$. As before, we are concerned with tuning the threshold λ_θ so the long term reward rate

$$R^{\lambda_\theta} = \frac{p_H\bar{m}_H + p_L\bar{m}_L}{p_H\bar{T}_H + p_L\bar{T}_L + \tau}$$

is maximized. Unlike the non-depleting case, we cannot frame and explicitly solve a MFPT problem to determine the mean times spent \bar{T}_H and \bar{T}_L and number of food chunks consumed \bar{m}_H and \bar{m}_L in the high and low patches. However, we can compute these quantities via Monte Carlo sampling. There is an internal optimum $\lambda_\theta^{\text{opt}}$ that maximizes the reward rate R^{λ_θ} across the environment, which deviates only slightly when little food is available (Fig. 7A) and is well matched in the case of high food availability (Fig. 7B) to a forager that knows the patch type upon patch arrival (compare black/blue curves). This suggests the forager typically learns the patch type before departing in environments with sufficient food, so their departure strategy will converge to that of an observer who already knows the patch type.

We analyzed this trend by comparing departure times of observers that initially know their current patch type to those that do not (Fig. 7C). Indeed, foragers in sparser environments (lower average initial food amount $\bar{m}_0 = (\bar{m}_0^H + \bar{m}_0^L)/2$) stay in low patches too long (blue curve) and leave high patches too soon (red curve) in comparison to observers that immediately know their patch type. This is due to their uncertainty about their current patch before and at the time of departure (grey curve). On the

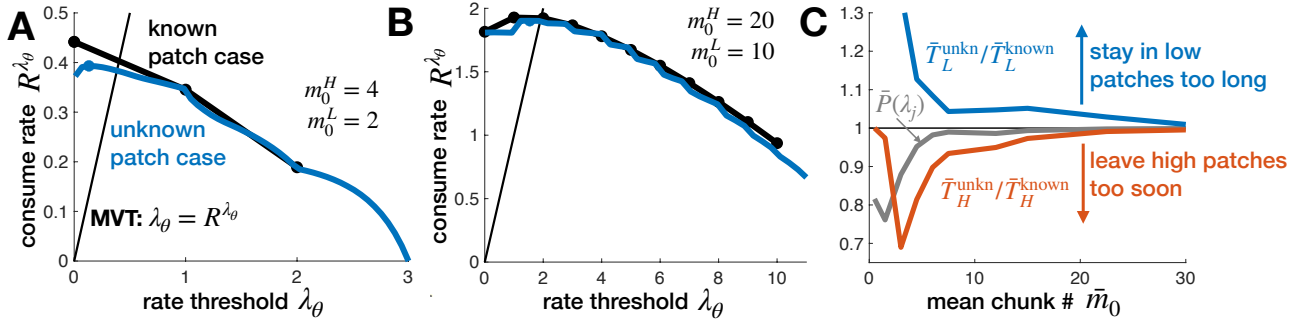


FIG. 7. Depleting environments with known versus unknown patch reward rates. In the ‘known’ case, the forager knows the patch type and initial reward rate upon arrival, whereas in the ‘unknown’ case the forager must infer the path type. Patch leaving strategies in the unknown case approach those of the known case in the high chunk count limit. **A.** Total reward rate R^{λ_θ} as a function of the estimated arrival rate λ_θ at which the observer departs a patch. For low initial food levels m_0^H and m_0^L , R^{λ_θ} near the optimum (dots) deviates between the case of initially known patch type (black) vs. initially unknown (blue). Both thresholds are less than the MVT prediction due to discreteness of chunk arrivals. **B.** Reward rates of the known/unknown cases converge at higher initial food levels. **C.** Deviations in optimal reward rate between known and unknown cases are due to less certainty $\bar{P}(\lambda_j)$ (grey line) about patch yield rate upon departure, leading the observer to stay in low patches too long (blue line) and leave high patches too soon (red line). Other parameters are $\rho = 1$, $\tau = 5$, and $p_H = 0.5$. Mean rates and departure times in the unknown patch case were computed from 10^6 Monte Carlo simulations.

other hand, in more plentiful environments, foragers accumulate enough evidence about the current patch that their departure times resemble those of observers that know their patch type, recapitulating the MVT.

Uncertainty thus drives deviations from the MVT. Foragers that exit patches before fully learning their type will typically understay (overstay) high (low) patches. Key to these uncertainty-driven decisions is an environment in which high/low patches are different enough to warrant different strategies, but similar enough as to not be immediately distinguishable. We now extend our analysis to the case of environments small enough so that foragers eventually return to previously depleted patches.

C. Returning to a previously depleted patch

Assuming foragers do not return to previously depleted patches is a reasonable approximation in large environments, since there is a low probability of the observer returning to a previously visited patch if drawing randomly from their environment. However, in small environments, past patch depletion impacts the value of future patch visits sooner (Fig. 8A). We briefly analyze this case now, identifying strategies that minimize the time to fully deplete the environment before departing for another environment.

Single chunk-two patch environment. First, consider the simplest nontrivial case, a two patch environment in which there is a single chunk of food ($m_0^H = 1$ so $\lambda_H = \rho$ initially and $m_0^L = 0$). When the forager departs one patch, they journey to the other patch with transit time τ . We seek a waiting time strategy similar to that developed for depleting patches (Fig. 8B). The observer waits a time T_θ in their current patch before leaving to

the other patch (if they do not encounter a food chunk). Note, this differs from thresholding the LLR ($y = \theta$) since the first patch visit would be shorter in that case, but this approach allows an explicit approximation of the optimal strategy. If the forager is in the high patch, they encounter the food chunk with probability $1 - \pi_\theta = 1 - e^{-\lambda_H T_\theta}$. In the low patch, they always depart after a time T_θ . Thus, the mean time to encounter the chunk in the environment is given by: (a) the time spent in the high patch visit on which the chunk is discovered plus (b) the initial time spent in the low patch when that is the first patch visited (scaled by that probability 1/2) plus (c) the other time spent visiting patches and not finding the chunk:

$$\begin{aligned} \bar{T}_{\text{find}}(T_\theta; \lambda_H) &= \frac{1 - (\lambda_H T_\theta + 1)e^{-\lambda_H T_\theta}}{\lambda_H(1 - e^{-\lambda_H T_\theta})} + \frac{T_\theta + \tau}{2} \\ &\quad + 2(T_\theta + \tau)(1 - \pi_\theta) \sum_{j=0}^{\infty} j \pi_\theta^j \\ &= \frac{\lambda_H T_\theta + 3\lambda_H \tau - 2 + (2 + \lambda_H T_\theta + \lambda_H \tau)e^{\lambda_H T_\theta}}{2\lambda_H(e^{\lambda_H T_\theta} - 1)}. \end{aligned} \quad (28)$$

This is minimized by taking $T_\theta = T_\theta^{\text{opt}}$ such that

$$e^{2\lambda_H T_\theta^{\text{opt}}} - 2(\lambda_H T_\theta^{\text{opt}} + 2\lambda_H \tau)e^{\lambda_H T_\theta^{\text{opt}}} - 1 = 0,$$

For large $e^{\lambda_H T_\theta^{\text{opt}}}$, we approximate the critical point equation as $e^{\lambda_H T_\theta^{\text{opt}}} \approx 2\lambda_H T_\theta^{\text{opt}} + 4\lambda_H \tau$, which has a compact asymptotic expansion:

$$\begin{aligned} T_\theta^{\text{opt}} &\approx \frac{1}{2\lambda_H} \left[-2W_{-1} \left(-\frac{e^{-2\lambda_H \tau}}{2} \right) - 4\lambda_H \tau \right] \\ &\approx \frac{1}{\lambda_H} \log(4\lambda_H \tau + 2 \log 2), \end{aligned} \quad (29)$$

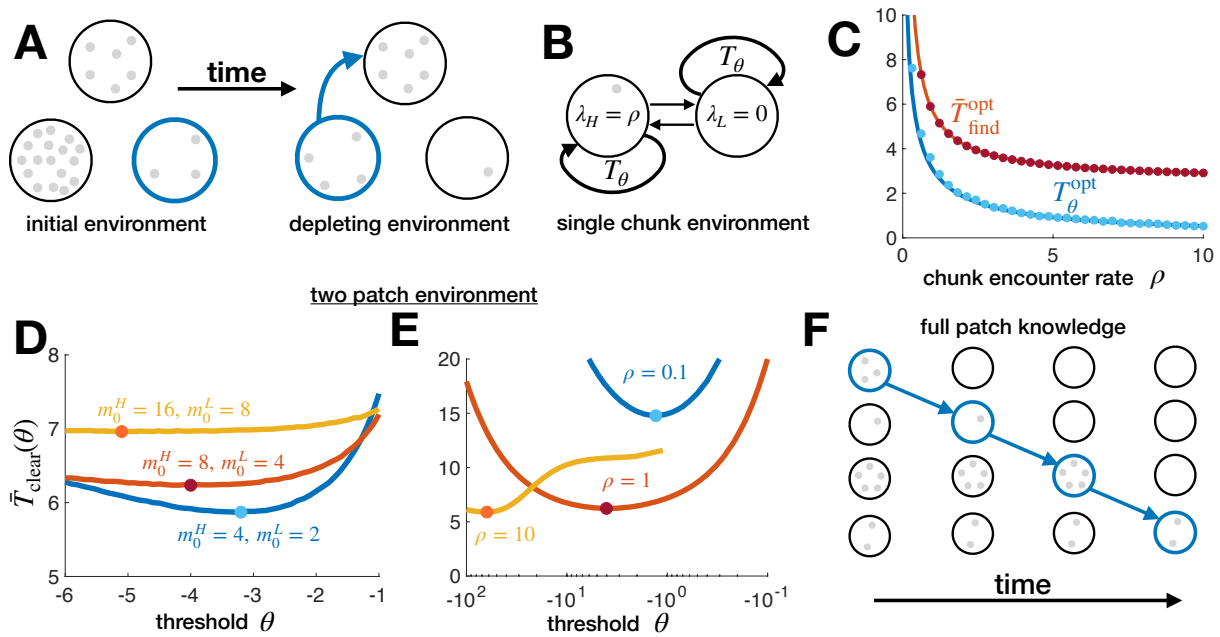


FIG. 8. Returning to patches in fully depleting environments. **A.** Environmental food depletion in small environments where previously visited patches (blue) remain depleted indefinitely. **B.** Strategy in a two patch environment with a single food chunk: Switch patches after a finite search time T_θ . **C.** Optimal switch time T_θ^{opt} (approximated by Eq. (29)) and the total mean time to find a chunk $\bar{T}_{\text{find}}^{\text{opt}}$ (Eq. (28)) decrease as the rate of chunk detection ρ increases. Theory compares well with averages from 10^6 Monte Carlo simulations (dots). **D.** In binary environments, the optimal LLR threshold θ^{opt} (dots) for minimizing the time to clear the environment decreases as the amount of food in the environment increases. $\rho = 1$ is used here. **E.** Optimal threshold θ^{opt} decreases as the rate of encountering a chunk increases: $m_0^H = 8$, $m_0^L = 4$. $p_H = 0.5$ and 10^6 Monte Carlo simulations for each curve. **F.** Optimal strategy given full knowledge of chunk count in each patch is to clear each patch sequentially.

where W_{-1} is the -1 branch of the Lambert W function. This shows that the optimal waiting time scales inversely with the initial arrival rate λ_H of the high patch. The minimal time to find the chunk is approximated by substituting Eq. (29) into Eq. (28). Both quantities decrease with ρ (Fig. 8C), owing to the fact that environments with higher discovery rates $\lambda_H = \rho$ require less time in the high patch for the observer to find the food chunk. When the detection rate ρ is small, the optimal switch time T_θ^{opt} (i.e. the decision strategy) is more sensitive to small changes in ρ .

General binary environment. A similar approach can be used in binary environments with an arbitrary initial number $m_0^H > m_0^L$ of food chunks in the high and low patch. To minimize the number of required patch switches, the forager should first clear m_0^L chunks of food from the starting patch and then wait until the observer's LLR $y = \log \frac{P(\lambda_H(t)|x(t))}{P(\lambda_L(t)|x(t))}$ (computed from Eq. (4)) falls below the threshold $\theta < \log \frac{p_H}{1-p_H}$. After the first couple of patch switches, the time spent in a patch if food is not discovered will be $T_\theta = -2\theta / [\rho(m_0^H - m_0^L)]$. While an explicit formula for $\bar{T}_{\text{clear}}(\theta)$ can be obtained by marginalization, the resulting expressions are lengthy and do not provide insight into the trends underlying the optimal strategy. Thus, we use Monte Carlo sampling to determine how the optimal θ varies with the environmental

parameters.

As the amount of food in the environment increases, the time to clear it increases, but only slightly (Fig. 8D). Since the observer has more evidence about the kind of patch they are in after consuming m_0^L chunks as m_0^L grows, they can more quickly find the $(m_0^L + 1)$ th chunk. Moreover, the patches become more discriminable as the difference $m_0^H - m_0^L$ increases, such that the forager can afford more certainty before switching patches, meaning that the optimal θ^{opt} is lower. Increased discriminability also results from increasing the rate of chunk encounters ρ (Fig. 8E). With increased discriminability, both the optimal threshold θ^{opt} and mean time to clear the environment $\bar{T}_{\text{clear}}^{\text{opt}}$ decrease owing to the forager's ability to discover chunks more quickly.

More patches. We expect it is possible to extend the above strategy to the case of N environmental patches. However, the dimensionality of the optimization problem will grow quickly, so we save a detailed analysis of this case for future work.

Nonetheless, we can easily identify a strategy for minimizing the time to clear the environment in the case of N patches assuming the observer knows the patch type upon arrival. If patch j has m_0^j chunks of food initially, the forager should still aim to minimize the total time spent in transit between patches. To do this,

they should fully clear each patch before moving on to the next (Fig. 8F). Assuming the forager always chooses an unexplored patch when departing an emptied one, the minimal mean time to clear the environment of food will be

$$\bar{T}_{\text{clear}}^{\text{opt}} = (N-1)\tau + \frac{1}{\rho} \sum_{j=1}^N H_{m_0^j} = N[\tau + \bar{H}/\rho] - \tau,$$

where the time to clear the j th patch is $H_{m_0^j}/\rho$ as in Eq. (20) for homogeneous depleting environments and \bar{H}/ρ is the mean clearance time of patches across the environment. Thus, the time to clear the environment scales linearly with the number of patches N (assuming the mean patch clearance time is unchanged by N).

Were the forager to use another strategy, they would return to each patch multiple times to clear them. This would mean the transit contribution would be $\bar{T}_{\text{transit}} > (N-1)\tau$, and the total time to clear patch j would be no less than $H_{m_0^j}/\rho$, so $\bar{T}_{\text{clear}} > \bar{T}_{\text{clear}}^{\text{opt}}$.

There are a number of extensions to this case of a fully depleting environment. Since the observer must consider both transit times between patches and patch yields that depend on visit history, this is a qualitatively different problem from the multi-armed bandit (MAB) problem. Typically, bandit problems do not involve depletion or action-dependent changes in the yield of arms [59, 60]. Our setup thus provides a rich class of optimal switching strategy problems which warrant future investigation.

V. LEARNING THE DISTRIBUTION OF PATCH TYPES

Our analysis so far has assumed the forager has complete knowledge of the distribution of patch arrival rates in the environment. However, typically animals arrive in a new environment with uncertainty about the quality of patches they will encounter [28, 61–63]. Thus, animals should use information from patch encounters to estimate statistics of patch yields in the environment. In our sequential updating model, this can be implemented with a higher level of inference in the model (Fig. 1B) whereby the observer refines a prior over the initial arrival rate λ_0 using the past N patch visits $p(\lambda_0|x_{1:N}(t))$. We assume each new patch a forager enters has not been previously depleted.

Here, we measure learning performance using the time it takes to refine the mean squared error (MSE) to some threshold. Interestingly, we find that foragers in depleting environments learn the initial rate of arrival faster than those in non-depleting environments.

A. Homogeneous environment

A homogeneous environment is described by a single parameter λ_0 , the initial arrival rate of food within all

patches. We will start by considering the simple non-depletion case, and then move to consider how learning occurs in depleting environments. This will illuminate the differences between learning in these two cases before moving to binary environments.

Non-depleting environments. In a non-depleting environment, an agent intending to maximize the speed they learn the patch arrival rate of food λ_0 should remain in a single patch indefinitely. In this case, the rate learning process is equivalent to learning the rate of a sequence of exponentially distributed waiting times. Given a sequence of food encounters $x(t) = \sum_{j=1}^{K(t)} \delta(t - t_j)$, the observer uses the time t , count $K(t)$, and initial prior $p_0(\lambda_0)$ to estimate the rate λ_0 so Bayes rule states

$$p(\lambda_0|x(t)) = \frac{1}{p(K(t))} p(K(t)|\lambda_0) p_0(\lambda_0) \propto (\lambda_0 t)^{K(t)} e^{-\lambda_0 t} p_0(\lambda_0), \quad (30)$$

and the maximum likelihood estimate (MLE) of the rate λ_0^* is given by the implicit equation

$$t\lambda_0^* = K(t) + \lambda_0^* p'_0(\lambda_0^*)/p_0(\lambda_0^*).$$

For a flat, improper prior $p_0(\lambda_0) \equiv 1$, the MLE is $\lambda_0^* = K(t)/t$, and for an exponential prior $p_0(\lambda_0) = ae^{-a\lambda_0}$, the MLE is $\lambda_0^* = K(t)/(t+a)$. In either case, the MLE is consistent in the limit as $t \rightarrow \infty$:

$$\lim_{t \rightarrow \infty} \lambda_0^* = \lim_{t \rightarrow \infty} \frac{K(t)}{t} + \lim_{t \rightarrow \infty} \frac{\lambda_0^* p'_0(\lambda_0^*)/p_0(\lambda_0^*)}{t} = \lambda_0^{\text{true}}.$$

Each food encounter tends to narrow the distribution $p(\lambda_0|x(t))$ and the MLE tends to move toward the true arrival rate λ_0^{true} (Fig. 9A). One measure of learning is the rate of decrease of the normalized MSE. Since the estimator is consistent, we expect

$$\text{MSE}(t; \lambda_0^{\text{true}}) = \int_0^\infty \frac{(\lambda_0 - \lambda_0^{\text{true}})^2}{(\lambda_0^{\text{true}})^2} p(\lambda_0|x(t)) d\lambda_0$$

to decrease on average over time. That is, $\frac{d}{dt} \bar{\text{MSE}}(\lambda_0^{\text{true}}, t) < 0$ where

$$\bar{\text{MSE}}(t; \lambda_0^{\text{true}}) = \int_0^\infty \frac{(\lambda_0 - \lambda_0^{\text{true}})^2}{(\lambda_0^{\text{true}})^2} \bar{p}(\lambda_0|x(t)) d\lambda_0, \quad (31)$$

and where the average is over realizations of arrival times $p(K(t)|\lambda_0^{\text{true}})$, leading to

$$\bar{p}(\lambda_0|x(t)) = \sum_{K=0}^\infty \frac{(\lambda_0^{\text{true}} t)^K (\lambda_0 t)^K e^{-\lambda_0^{\text{true}} t} e^{-\lambda_0 t} p_0(\lambda_0)}{(K!)^2 p(K)}.$$

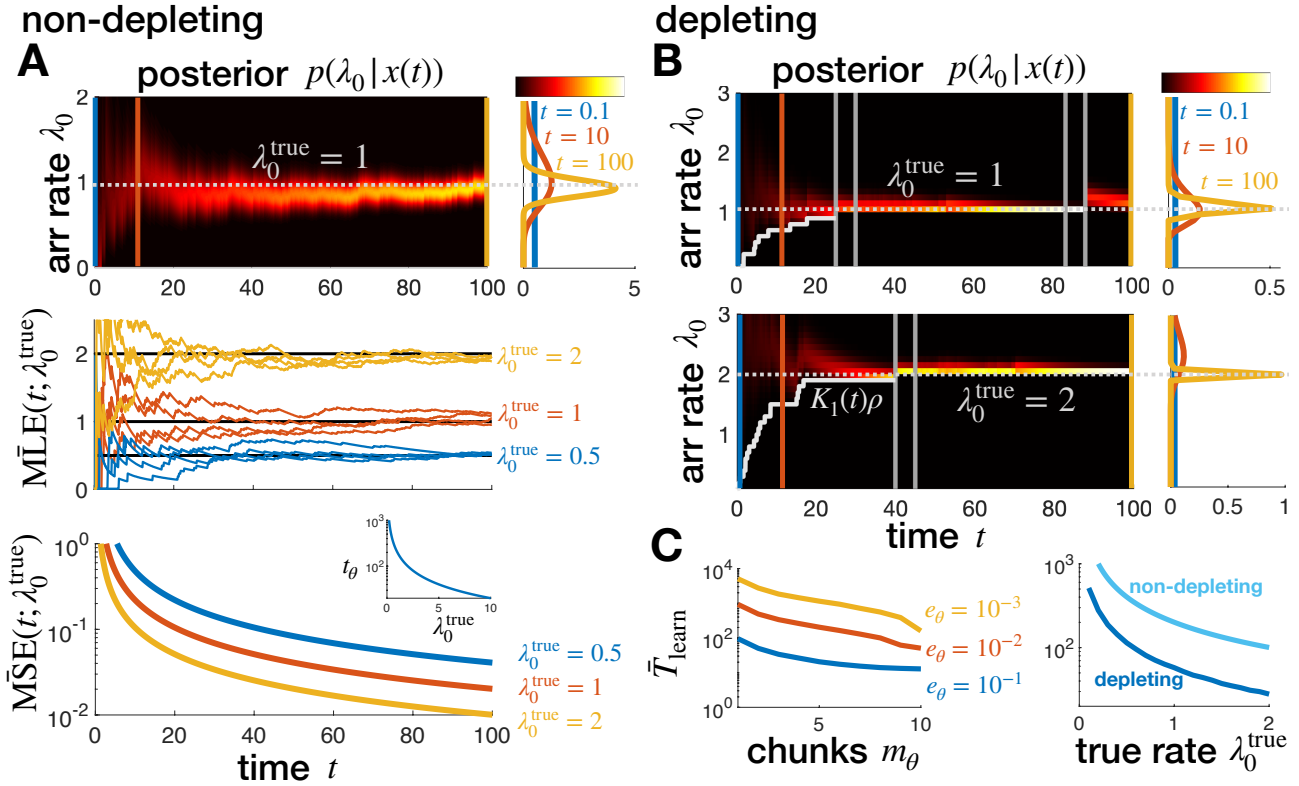
For a flat, improper prior $p_0(\lambda_0) \equiv 1$,

$$p(\lambda_0|K(t)) = \frac{t}{K!} (\lambda_0^t)^K e^{-\lambda_0 t},$$

so

$$\bar{p}(\lambda_0|x(t)) = t e^{-(\lambda_0^{\text{true}} + \lambda_0)t} \sum_{K=0}^\infty \frac{(\lambda_0^{\text{true}} \lambda_0 t^2)^K}{(K!)^2}$$

homogeneous environments



binary environments

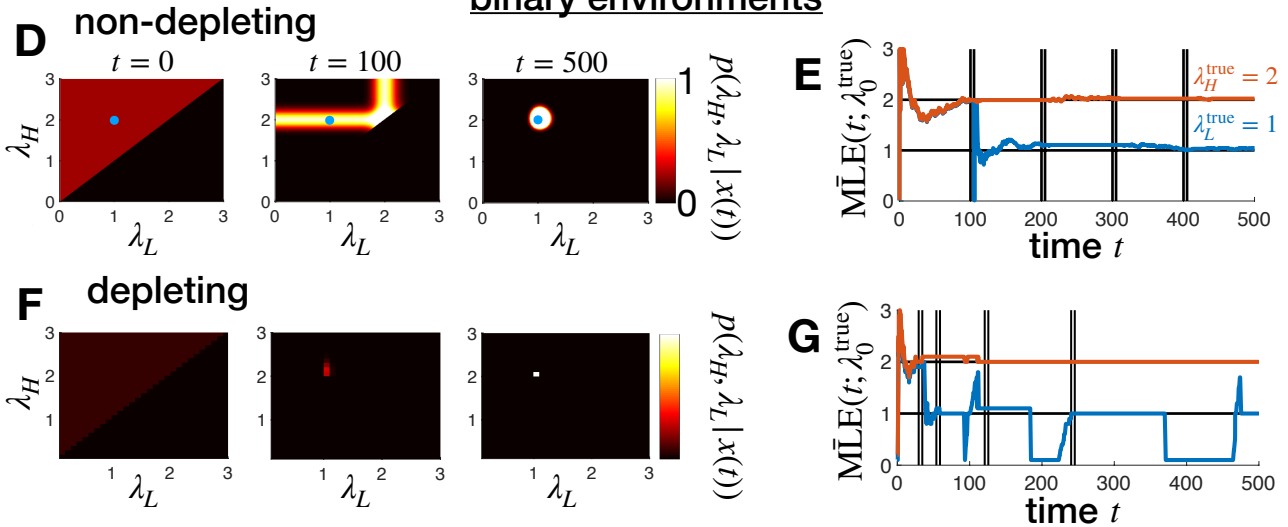


FIG. 9. **Learning reward arrival rates.** Patch arrival rates are learned more rapidly in depleting than non-depleting environments. **A.** Posterior pdf $p(\lambda_0|x(t))$ is refined leading to a maximum likelihood estimate (MLE) which approaches the true arrival rate: results for a non-depleting environment. Average mean squared error (MSE), given by Eq. (31), decreases more rapidly for higher λ_0^{true} . **B.** Posterior pmf $p(\lambda_0|x(t))$ peaks more rapidly in depleting environments due to exclusion of rates $\lambda_0 < K(t)\rho$. **C.** Mean time to learn \bar{T}_{learn} to within $\text{MSE} < e_\theta$ is smallest when chunks consumed equals the total number in the patch $m_\theta = m_0 = 10$ (left), so the best policy (right) is to exit when MLE of total number of chunks is consumed ($m_\theta^{\text{opt}} = \text{MLE}(m_0)$). Here $\rho = 0.1$. Note lower \bar{T}_{learn} than in non-depleting case. **D.** In binary environments, posterior pdf $p(\lambda_H, \lambda_L|x(t))$ is first refined over the first patch type visited ($\lambda_H = 2$ here), then over the second patch type visited ($\lambda_L = 1$), and eventually reaches a peak near true (λ_H, λ_L) . **E.** Thus, eventually an accurate MLE is obtained for both patch types. Departures every $t = 100$ time units. **F,G.** Learning is more rapid in depleting environments, because low arrival rates are ruled out. Patches departed after MLE chunks consumed.

$$\begin{aligned}
 &= tI_0(2t\sqrt{\lambda_0\lambda_0^{\text{true}}})e^{-(\lambda_0^{\text{true}}+\lambda_0)t} \quad (32) \\
 &\equiv F(t, \lambda_0, \lambda_0^{\text{true}}),
 \end{aligned}$$

where $I_0(z)$ is the zeroth order modified Bessel function of the first kind. Eq. (32) can thus be substituted into Eq. (31) and integrated using quadrature to calculate the ensemble mean of the MSE as a function of time:

$$\begin{aligned}
 \text{M}\bar{\text{S}}\text{E}(t; \lambda_0^{\text{true}}) &= \int_0^\infty \frac{(\lambda_0 - \lambda_0^{\text{true}})^2}{(\lambda_0^{\text{true}})^2} F(t, \lambda_0, \lambda_0^{\text{true}}) d\lambda_0 \\
 &= \int_0^\infty (x - 1)^2 F(s, x, 1) dx,
 \end{aligned}$$

where we have made the changes of variables $\lambda_0^{\text{true}}x = \lambda_0$ and $s = \lambda_0^{\text{true}}t$. This reveals that the rate of decay of the normalized mean MSE increases linearly in λ_0^{true} , such that the time t_θ to obtain a threshold level of mean accuracy scales inversely with λ_0^{true} (Fig. 9A bottom inset). Thus, the observer learns the arrival rate of the environment more quickly in higher rate environments.

Note, there are a number of ways to modify this basic setup. First, the forager may have a more informed prior over λ_0 based on experience with previous environments. In this case the updating process is the same and only the prior $p_0(\lambda_0)$ changes. Had we considered an observer that moves between patches, the only difference would be that the observer gains no information about the arrival rate while in transit but carries their posterior forward upon arrival in the new patch. We now extend this framework to consider depleting environments.

Depleting environments. In depleting environments, the forager's belief update changes in two key ways. First, to continue learning, they must eventually depart their current patch and continue observing in a new patch whose arrival rate is reset to the true initial arrival rate λ_0 . Second, each chunk encounter reduces the arrival rate within the patch, so the space of possible arrival rates must be adjusted. In particular, all arrival rates $\lambda_0 < K(t)\rho$ are eliminated from the posterior, as they are impossible given $K(t)$ chunks have arrived. With these facts in mind, we examine how the time to learn the arrival rate depends on patch exit strategy and true arrival rate. Following depleting foraging strategies, we assume the departure threshold on the number of chunks consumed depends on the MLE of the arrival rate.

To determine an efficient strategy for inferring the arrival rate λ_0 , we first determine the optimal departure chunk count given a specific true λ_0^{true} . We then examine how an observer who chooses this chunk count threshold based on their MLE limits their time to obtain a given desired MSE. As before, the posterior over the initial arrival rate λ_0 can be calculated using Bayes rule, so within a single patch

$$p(\lambda_0|x(t)) = \frac{1}{p(K(t))} p(K(t)|\lambda_0) p_0(\lambda_0) \propto \frac{m_0! e^{-\lambda_0 t} p_0(\lambda_0)}{(m_0 - K(t))!},$$

for $\lambda_0 \geq K(t)\rho$, where $m_0 = \lambda_0/\rho$ is the initial count associated with the initial rate λ_0 . Each food encounter

generally narrows the distribution $p(\lambda_0|x(t))$ and decreases the number of chunks in the patch. Given N patch visits, the forager can accumulate evidence from each visit as a separate term in the posterior product:

$$p(\lambda_0|x(t)) \propto \prod_{j=1}^N \frac{m_0! e^{-\lambda_0 t_j}}{(m_0 - K_j(t_j))!} \cdot p_0(\lambda_0), \quad (33)$$

where t_j is the time spent in the j th patch and $K_j(t_j)$ is the number of chunks encountered (Fig. 9B top).

Averages from simulations show the best strategy for minimizing the time to learn the true rate λ_0^{true} is to depart the patch when all the chunks predicted by the observer's current MLE are consumed (Fig. 9C left). We see evidence of this, for instance, when attempting to infer $\lambda_0^{\text{true}} = 1$ (when $\rho = 0.1$ and $m_0 = 10$). The forager's best strategy for minimizing the time to learn this rate is to consume all 10 chunks of food in a patch before departing. Thus, an effective strategy for rapidly learning the initial arrival rate λ_0 across environments is to leave a patch once the number of chunks consumed is greater than or equal to the MLE for the initial chunk count m_0 ($m_\theta^{\text{opt}} = \text{MLE}(m_0)$). Using this strategy, the time to learn the rate decreases as the true rate increases (Fig. 9C right) in line with findings in the non-depleting case, but much faster due to exclusion effects of depletion.

B. Binary environments

A forager learning two patch arrival rates $\lambda_H > \lambda_L$ clearly must visit more than one patch, and each successive patch visit increases the likelihood that both patch types have been visited. As before, we start by considering the non-depleting environment in which learning is slower, and then move to studying the inference process in a depleting environment.

Non-depleting environments. In a single patch (j), given a flat prior ($p(\lambda_0)$ constant), a sequence of food encounters $x_j(t)$ yields the posterior

$$p(\lambda_0|x_j) \propto (\lambda_0 t_j)^{K_j(t_j)} e^{-\lambda_0 t_j} \equiv Q(\lambda_0; t_j, K_j),$$

over the possible arrival rate λ_0 in patch j , where t_j is the time spent and $K_j(t_j)$ is the number of chunks encountered in patch j (as in Eq. (30)).

Given N patch visits and a flat prior ($p(\lambda_H, \lambda_L)$ constant) over the rates, the forager combines information across patches by conditioning on the probability they are in a high/low patch. The posterior for the arrival rates λ_k ($k = H, L$) given patch visits $j = 1, \dots, N$ is then:

$$\begin{aligned}
 p(\lambda_H, \lambda_L|x(t)) &\propto \prod_{j=1}^N p(x_j(t_j)|\lambda_H, \lambda_L), \quad \lambda_H > \lambda_L, \\
 &\propto \prod_{j=1}^N [p_H Q(\lambda_H; t_j, K_j) + p_L Q(\lambda_L; t_j, K_j)],
 \end{aligned}$$

where $p_k \equiv p(\lambda_j = \lambda_k)$ is assumed known to the observer.

The posterior updating is thus refined based on the arrival rate of the current patch (about λ_H^{true} when in a high patch and about λ_L^{true} when in a low patch: Fig. 9D). This process continues as the agent visits each type of patch, as seen in time series of the MLE, whereby the estimate of the arrival rate for the patch not being visited remains fairly constant, while that of the visited patch changes (Fig. 9E).

Depleting environments. Lastly, we derive the inference strategy for learning in binary environments with depleting patches. Learning the arrival rate of a single patch is given by Eq. (33), and using a flat prior we have

$$p(\lambda_0|x_j(t)) \propto \frac{(\lambda_0/\rho)!}{(\lambda_0/\rho - K_j)!} e^{-\lambda_0 t_j} \equiv Q_\rho^j(\lambda_0),$$

where t_j is time spent and K_j is the number of chunks encountered in patch j . Conditioning on the patch type being visited yields the posterior for the initial arrival rates λ_0^k ($k = H, L$) given patch visits $j = 1, 2, \dots, N$:

$$\begin{aligned} p(\lambda_0^H, \lambda_0^L|x(t)) &\propto \prod_{j=1}^N p(x_j(t_j)|\lambda_0^H, \lambda_0^L), \quad \lambda_0^H > \lambda_0^L, \\ &\propto \prod_{j=1}^N [p_H Q_\rho^j(\lambda_H) + p_L Q_\rho^j(\lambda_L)], \end{aligned}$$

where $p_k \equiv p(\lambda_0^j = \lambda_0^k)$.

As in the non-depleting environment, the posterior is first refined over the first patch type visited ($\lambda_H = 2$ in Fig. 9F). However, as opposed to the homogeneous environment, low rates $\lambda_L < K\rho$ are never entirely ruled out, since there is always a possibility the forager is in the high patch. As such, the MLE for the high patch is refined more quickly while the low patch MLE continues to jump around (Fig. 9G).

We could also consider models that learn the depleting increment ρ or distribution of transit times τ using a similar Bayesian sequential updating framework. Presumably, natural foragers would learn environmental parameters through related evidence accumulation processes. Our work here demonstrates that sequential Bayesian inference affords a powerful framework for understanding how a forager could learn hierarchically using a fairly compact model. A key finding of the above analysis is that depletion itself provides additional information about the arrival rate, allowing the observer to refine their estimate more rapidly than in non-depleting environments.

VI. DISCUSSION

A. Summary

Patch leaving decisions are an essential component of foraging. Our model uses principles of probabilistic inference to establish a normative theory of patch leaving

decisions, as well as a framework for learning about resource availability in the environment.

The model yields several key takeaways concerning optimal decision strategies in a variety of patch foraging contexts (see Fig. 10 for an overview and Table I for details). In the idealized case in which foraging does not deplete patches, the optimal strategy is for a forager to minimize their time to find and remain in the highest yielding patch in the environment. This is accomplished by triggering patch leaving decisions when the log-likelihood ratio for the probability of high return versus other patches falls below a threshold. When a forager depletes patches, we found that across a broad range of environmental parameters the best strategy for maximizing the environmental reward rate is to depart a patch when the in-patch reward rate matches the average return rate of the environment, i.e. corresponding to the marginal value theorem (MVT). However, if there is high uncertainty about the patch type, the forager will stay longer in low yield patches and shorter in high yield patches than predicted by the MVT. When a forager fully depletes their environment, an optimal strategy minimizes the number of transits between patches in order to deplete the environment as quickly as possible. Lastly, we found an observer that learns the rate of food arrival within patches learns more quickly in depleting than non-depleting environments.

Throughout the study, we compute and vary metrics that have parallels used in systems neuroscience such as reaction times and discriminability. We solved the model equations via a combination of first passage time methods for stochastic processes along with Monte Carlo sampling of jump processes corresponding to foragers' beliefs. Since the contexts analyzed in our study (Fig. 10; Table I) can be realized experimentally, our model provides a framework for the formal quantitative analysis of behavior in patch foraging experiments.

B. Relation to other work

Animals often approximate Bayesian foraging. We have treated patch leaving as a statistical inference problem within a Bayesian framework, assuming the animal has knowledge of the transit time between patches, the impact of food depletion on the food arrival rate, and (in the case without learning) the distribution of patch arrival rates. Our theoretical treatment of patch leaving decisions builds on previous Bayesian models of foraging [34, 36–42], experimental studies that ask if animals behave as Bayesians [43–49], and other experiments that show how an animal's prior knowledge about an environment modulates its foraging behavior [64–66]. For example, bumblebees [32] and inca doves [67] adjust their foraging strategies in response to the predictability of the environment, as a Bayesian forager would, but other work shows this is not a universal trend [68]. Patch leaving decisions may deviate from Bayes optimality due to ani-

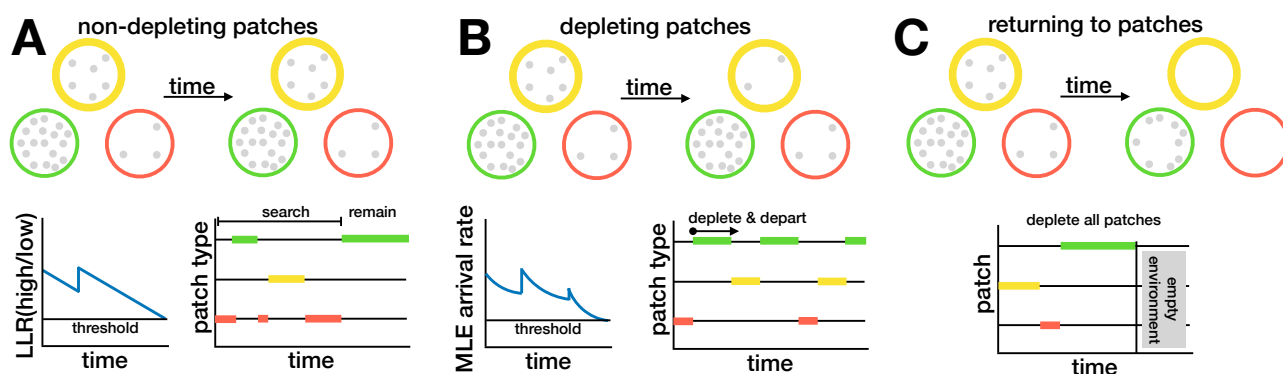


FIG. 10. **Summarized taxonomy of foraging strategies.** See Table I for details. **A.** In non-depleting environments, an ideal forager searches patches until finding and remaining in a high yielding patch. **B.** In environments with depleting patches, an ideal forager depletes a patch and departs when reward arrival rate reaches a threshold. **C.** In fully depleting environments, an ideal forager should fully deplete patches before departing in order to minimize the time to clear the environment.

mals becoming risk-averse in variable environments [69]. Apart from the behavioral significance of priors in guiding animals' foraging behavior, such cases can help us understand the types of probabilistic neural computations performed in uncertain and dynamic environments.

Inferential decisions can validate the marginal value theorem. The marginal value theorem (MVT) [22] provides a baseline patch departure rule for comparing optimal foraging to ecological behavior [12]. However, the MVT is limited in its applicability, as it classically describes the case of continuous rewards where the forager has perfect knowledge of resource availability, and does not provide a mechanistic account of how a forager reaches a decision to leave a patch. To address these limitations, Davidson & El Hady [31] developed a generalized mechanistic model of patch leaving decisions with parameters representing the discreteness of reward arrival, belief noise, and environmental averaging. In this model, patch leaving decisions are due to a threshold crossing process by an accumulator variable. Parameters can be adjusted to represent a wide range of foraging strategies not limited to the MVT, and specifically addressing environmental uncertainty. This model inspired our current work, which demonstrates that optimal patch leaving decisions involve evidence accumulation. We have identified uncertainty in the current reward rate as a key driver of deviations from the MVT. Previous studies modeling patch departure under discrete rewards [25] have formulated mechanistic descriptions of departure decisions [20, 21, 26, 29], but none have analyzed such a broad range of task conditions to provide a comparative study of strategies.

Patch foraging as modified multi-armed bandit. Patch leaving behavior involves deciding when to leave the current patch being harvested, noting that available rewards in the next patch may be unknown; in the simplest case departures lead to a random sample of the subsequent patch from the environment. In contrast, the multi-armed-bandit (MAB) task [70] assumes an agent has some control over which patch (arm) is chosen next,

often does not consider patch depletion, and classically ignores switching costs. However, patch leaving and the MAB task can become similar in certain limits. For instance, patch leaving that involves no depletion, zero switching costs, and memory of past locations is somewhat similar to the classic MAB.

To illustrate the differences, consider an environment with two patch types H and L , no depletion of patches, no switching cost, and where the forager knows the patch reward rates values λ_H and λ_L (treated in Figs. 2 & 3). Within a patch, the forager uses its experience to decide when to leave the current patch, since they may not know the patch type on arrival. If the forager does not retain memory of the locations of specific patches, then they will randomly come upon the next patch of either type H or L according to the statistics of the environment, and must repeat the process of inferring the reward arrival rate. Now consider a K -arm bandit where some proportion have reward probability λ_H and some have λ_L . At each step, the gambler chooses which of the K arms to exploit, and continually updates its estimate of which patches are H versus L , as the space of arms is typically small enough to learn this well. Even with the assumptions of no depletion and zero switching costs, as formulated these are still different decision problems: to formulate the MAB problem as a patch foraging problem, the gambler would have a choice at each step between the same arm or a randomly chosen different arm.

Regardless, considerations of switching costs, depletion, and uncertain future rewards are central to patch foraging and stay-or-go decisions, and related work with the MAB has considered some of these cases. Switching costs were considered by [60, 71, 72], and including these who generally led to remaining at the same arm for longer, similar to patch foraging. The problem of depleting rewards is related to the non-stationary bandit, in the specific case where arm reward rates are choice dependent [73, 74]. In a patch with depleting rewards, the expected return changes in a specific and predictable way, and therefore can be treated mathematically, as we

<i>Environment</i>	<i>Decision strategy/solution</i>	<i>Dependencies/Observations</i>
Non-depleting patches		
Objective: Minimize time to arrive and remain in highest yielding patch. Known: Food arrival rates of each patch type.		
<i>N</i> -patch types	Depart patch when likelihood of being in highest yielding patch falls below a threshold. $N = 2$: See Eqs. (3) & (11) and Fig. 2B; $N = 3$: See Eqs. (14) & (15) and Fig. 4B; $N > 3$: See Eqs. (17) and Fig. 5B.	<ul style="list-style-type: none"> Optimal strategy and arrival time depend on fraction of low versus high yield patches and discriminability between 1st/2nd highest yield patches (Figs. 3-5).
Continuum of patch types	Categorize patches as high or low yielding and depart patch if likelihood of a high patch is below threshold (See Eq. (19) and Fig. 5D).	<ul style="list-style-type: none"> Time to arrive and remain in high yielding patch is non-monotonic in departure threshold, and much longer when high patches are rare.
Depleting		
Objective: Maximize mean food intake rate R over a long time (several patches). Known: Initial yield rates of each patch type.		
1-patch type	Depart when arrival rate $\lambda(t)$ falls to a threshold value λ_θ .	<ul style="list-style-type: none"> Matches MVT (Eqs. (22) & (23) and Fig. 6B) except when there are very few chunks per patch (Fig. 6C) in which case the forager should empty the patch.
2-patch types: patch type known on arrival	Depart when current patch arrival rate $\lambda_j(t)$ reaches a threshold.	<ul style="list-style-type: none"> Represents the 'Depletion-dominated' regime, which recovers MVT (See Eqs. (25) & (26) and Fig. 6D).
2-patch types: empty low yield patch	Wait a time T_θ , then depart patch if no food is found. If encounter food at $t < T_\theta$, then use threshold on inferred yield rate to make leaving decision (similar to single patch type case)	<ul style="list-style-type: none"> 'Uncertainty-dominated' regime deviates from MVT. Optimal wait time and departure threshold λ_θ increase with prevalence of high yielding patch (Fig. 6E,F)
2-patch types: both high and low patches have food	Decision via threshold on current estimated yield rate $\tilde{\lambda}(t)$ (See Eq. (27)). Choose optimal threshold λ_θ that maximizes long term food intake rate.	<ul style="list-style-type: none"> Optimal return differs from known case when there are few food chunks per patch (Fig 7A), converges to known patch case when there are many chunks per patch (Fig 7B). Forager stays in low patches too long, leaves high patches too soon when there are few chunks per patch (uncertainty-dominated regime, Fig 7C).
Returning to patches		
Objective: Minimize time to harvest all food from the environment. Known: Yield rates of each patch type.		
Patch type known on arrival	Empty each patch completely before moving to the next one.	<ul style="list-style-type: none"> Time to clear environment increase linearly with number of patches in environment (Fig. 8F).
2-patches: 1 chunk	Leave current patch if food not found within waiting time of T_θ (search problem).	<ul style="list-style-type: none"> Optimal waiting time decreases with discovery rate ρ (Fig 8C) in contrast to case without returns (Fig. 6E).
2-patches: general case	Consume m_0^L chunks from first patch, then threshold belief to decide patch departure.	<ul style="list-style-type: none"> Time to clear environment depends weakly on chunk count (Fig. 8D) and decreases slightly as chunk encounter rate increases (Fig. 8E).

TABLE I. **Detailed taxonomy of patch foraging strategies.** Patch decision strategies depend on the environment, and trends in observables differ for the foraging environments considered in the model. Environments considered include non-depleting patches, depleting patches, and the case of returning to specific patches to fully deplete an environment (see also Fig 10). Columns describe important aspects of the optimal decision strategy for each case, along with key model results.

showed in Section IV. Limited/incomplete memory of the return of each arm is related to the non-stationary bandit, because non-stationarity of return means estimates of the return of previously visited arms slowly becomes more uncertain over time. Other MAB task designs consider changing conditions: for example, after an abrupt change in the environment (e.g. [75]), estimates of the return from each arm can be 'reset' [76]. A complete reset of the choice policy represents no memory of which arm to choose. This is related to the fact that a forager often has knowledge of only the reward rate of the current patch - finding the next patch requires exploration and possibly an uncertain amount of search costs. Within these contexts, patch foraging is fairly well described by a non-stationary bandit with (possibly uncertain) switching costs, limited memory or discounting of the choice policy of certain arms, and a specific form of non-stationarity. As such, we see our work providing use-

ful insights about efficient strategies for non-stationary MAB tasks.

Bayesian learning could be approximated by reinforcement learning. At best, animals approximate the optimal learning processes we have described mathematically [61, 77, 78]. One such approximation is the relative payoff sum (RPS)-learning rule whereby the probability of choosing a patch is proportional to the relative payoff that the respective patch has yielded so far [62], related to a matching law identified in operant conditioning [79]. A reinforcement learning (RL) framework can also be applied to patch leaving decisions, as discussed by Kolling and Adam [80]. These authors showed that while purely model-free RL fails to describe common aspects of patch leaving decisions, model-based predictions of expected rewards may be used to decide when to leave a patch, in conjunction with model-free RL that can be used to evaluate the values of different decision strategies across

patches.

In our theoretical treatment, we assume the forager uses the correct model of in-patch return rates, consisting of stochastic rewards with an underlying constant rate (non-depleting environment) or decreasing rate (depleting environment). Even if a Bayesian forager did not know a priori whether patches were depleting or not, or what the depletion increment was, a hierarchical Bayesian update could be used to learn this. In our work, the forager uses a threshold θ on the LLR to make patch leaving decisions in non-depleting environments, or a threshold λ_θ on the inferred arrival rate to trigger patch departures in depleting environments. Although we obtained explicit solutions for the optimal thresholds θ^{opt} or $\lambda_\theta^{\text{opt}}$ by considering values that lead to maximal return, we note that within the framework of [80], an agent could use model-free RL to learn the optimal values of these parameters. More work needs to be done in order to compare Bayesian mechanisms of learning to RL in the context of patch foraging decisions.

History and memory effects in patch foraging. Animals performing trained [81–83] and naturalistic behavior [84–86] make history-dependent decisions based on the stimuli they have experienced and the responses they have made. History-dependence is likely mediated by both short and long term memory processes. Memory for food availability is crucial for efficient foraging, for example in starlings [57, 87]. Movement and search patterns of foraging animals are also governed by environmental estimates from memory [88]. Our theoretical treatment mostly accounts for memory by assuming the forager retains knowledge of their patch’s statistics and the statistics of the environment, as agents update their patch type beliefs by sampling patch food resources.

C. Experimental applications: Behavior and neuroscience

Experimental realizations of foraging behavior.

There is currently a surge of interest in systems neuroscience to study naturalistic behavior as opposed to traditional experimental designs in which animals are trained over long times to perform particular behavioral tasks (‘trained behavior’). One important distinction between the type of foraging behavior we are studying here and trained decision making tasks in perceptual decision making is that the former is continuous, without a specific trial structure, while the latter have a (repeated) trial structure predetermined by experimenters. The theoretical framework in this work will help experimenters move from trial based behavior to continuous behavior without the loss of quantitative rigor. Studying behavior in this continuous paradigm offers important advantages over trial based paradigms as it allows multi-sensory integration to unfold, increases the fraction of time the animal is engaged in a task, engages the animal’s sensory-cognitive-motor loops activated by the environ-

ment’s natural statistics, and allows neural dynamics to unfold over multiple timescales [89].

Although recent work has examined foraging tasks within a systems neuroscience framework [90–93], there remain many open questions related to how the brain processes key aspects of foraging decisions. For example, experiments with the “Self-control preparation,” where an animal chooses between two choices, have significant behavioral differences with those that have a sequential foraging “patch preparation”, even though from an economic standpoint, the setups are equivalent [94]. Moreover, many tasks do not allow the animal to move extensively in space, which is a crucial aspect of foraging. For instance, when rats are required to physically move to perform foraging, the observed behavior differs from tasks that “simulate” foraging by presenting sequential choices or that consider visual search [95].

Neural implementation of patch foraging decisions. As we are aiming to provide a conceptual framework for patch foraging experiments in systems neuroscience, a crucial aim of such experiments is to map brain circuits underlying patch foraging decisions. The behavioral conceptual framework we have treated in this study includes behavioral mechanisms underlying patch leaving decisions in different types of environments. By extension, those behavioral mechanisms should map onto neural mechanisms. As in classic perceptual decision-making tasks, sequential updating models provide key insights into the types of neural computations one should expect to see in subjects performing those decisions [50].

Fitting our behavioral model to both the behavioral data and the neural data simultaneously [96] can unravel a series of intricate neural computations underlying particular behavioral strategies. To date, only a few brain areas have been shown to be involved in different aspects of foraging decisions: ventromedial prefrontal cortex encodes values of well-defined options, and anterior cingulate cortex encodes the average value of the foraging environment and cost of foraging [97] or signals changes in environmental context [98, 99]. Although experimentalists have traditionally targeted selected stereotypical areas, the increased availability and versatility of large scale recording techniques will enable future experiments to map the global circuits underlying foraging decisions, thus accessing a brain wide representation of intricate behavioral strategies [100].

Comparative patch foraging behavior. Comparative approaches, identifying similarities and differences across species, are a key aspect of the behavioral ecology of foraging [101–103]. Our model enables quantitative, model-based comparisons of decisions strategies between animals, which will improve our understanding of the aspects of evolution (e.g., selection pressure) shaping foraging behavior [11, 14]. Since neural circuits are shaped by these pressures too, this opens novel avenues for comparative systems neuroscience and an evolutionary understanding of neural computations during foraging.

D. Model extensions

Our model forms the foundation for multiple avenues of future theoretical study. One extension, discussed above, concerns using reinforcement learning to tune the patch leaving decision thresholds in the model. While we assumed the agent learns the qualities of different patch types, we fixed the transit time between patches; another extension could focus on learning and decision making in environments with distributed transit times. Such a feature could arise via spatial arrangements of patches, with potentially correlated resource availabilities across adjacent patches [40, 104].

Effective search is integral to survival in nature [105], and search behavior can give information on individual decision strategies. For example, a recent study found that rats can solve the stochastic travel-salesman problem using a nearest neighbour algorithm [106]. A similar approach could ask how an animal's search and navigation pattern interact with different patch leaving decisions to create an effective foraging strategy.

Our framework has also developed strategies assuming the forager utilizes constant thresholds on either the belief or estimated arrival rate. However, it is well known that in a variety of decision making contexts, optimal strategies can involve time-dependent decision thresholds [52, 53, 107, 108]. Typically, these results arise in the context of multi-trial experiments in which the quality of evidence on each trial varies stochastically and is initially unknown. In the non-depleting environments of our study, the quality of evidence is fixed across patch visits and beliefs are used to trigger departure decisions. As such, models in this context fit the assumptions of classic, constant threshold optimal policies. On the other hand, in uncertain binary depleting environments, we have projected a higher-dimensional description of the patch value to a single scalar estimate of the patch arrival rate. As such, we cannot be sure the parameterized model of patch departure decisions is optimal across all models. Leveraging methods from dynamic programming commonly used to set optimal decision policies would be a fruitful next step in ensuring the optimality of our patch leaving decision strategies.

We defined optimal patch foraging as maximizing resource intake. However, an alternative formulation could consider an explicit trade-off between gathering resources (exploitation) and gathering information about the environment (exploration). Information foraging frames an agent's movement strategy by prioritizing the value of information over resource gathering [109]. In our problem,

information foraging would target strategies that reduce the time to learn the environmental distribution of patch qualities (as in Section V), rather than maximize the reward rate, thus prioritizing the role of exploration and sampling in foraging [48, 110–113].

Although the natural world typically offers a continuous foraging space, considering discrete resource patches facilitates understanding by systematically dissecting fine scale dynamics of local search from global search [114, 115]. Additionally, animals seem to exploit local patches of food resources as if they had partitioned the world into patches. Nonetheless, not all resources exist in patches; a further extension of our theoretical framework could consider a continuous world without patches. Assuming some regularity in the distribution of food, it would then be efficient to consider stochastic versions of gradient ascent so the forager could orient itself in the direction of higher food concentrations.

Another important extension will be to consider interactions between agents, either through predator-prey interactions which affect foraging decisions [116], social foraging of groups collectively foraging [117, 118], or even competitive foraging [119]. In our model, the agent only receives direct (non-social) information about resource availability; in collective foraging, an individual receives both social and non-social information [120, 121]. This can significantly affect foraging decisions, for example when an individual must balance a desire to obtain more resources with a desire to maintain group cohesion.

To conclude, our model establishes a formal framework for a natural behavior (patch foraging) that can be studied in the same formal rigor as many trained behavioral tasks. Guided by the philosophy that formal tractability of a behavior entails neural mechanistic tractability, future work will build on this framework to generate testable hypotheses on the neural mechanistic underpinnings of foraging behavior.

ACKNOWLEDGMENTS

We thank Jan Drugowitsch for helpful discussions regarding the optimality of decision threshold policies across patch visits. ZPK was supported by a CRCNS grant (R01MH115557) and NSF (DMS-1853630). AEH acknowledges support by NIH grant 1R21MH121889-01. JDD acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2117 - 422037984, as well as support from the Heidelberg Academy of Science.

-
- [1] Adam J Calhoun, Sreekanth H Chalasani, and Tatyana O Sharpee. Maximally informative foraging by caenorhabditis elegans. *Elife*, 3, 2014.
- [2] Joshua S. Greene, Maximillian Brown, May Do-

bosiewicz, Itzel G. Ishida, Evan Z. Macosko, Xinxing Zhang, Rebecca A. Butcher, Devin J. Cline, Patrick T. McGrath, and Cornelia I. Bargmann. Balancing selection shapes density-dependent foraging behaviour. *Na-*

- ture, 539(7628):254–258, Nov 2016.
- [3] Marla B Sokolowski, Clement Kent, and Jonathan Wong. *Drosophila larval foraging behaviour: developmental stages*. *Animal behaviour*, 32(3):645–651, 1984.
- [4] Dan Landayan, David S Feldman, and Fred W Wolf. Satiating state-dependent dopaminergic control of foraging in drosophila. *Scientific reports*, 8(1):1–9, 2018.
- [5] Barnaby Marsh, Cynthia Schuck-Paim, and Alex Kacelnik. Energetic state during learning affects foraging choices in starlings. *Behavioral Ecology*, 15(3):396–399, 2004.
- [6] Cynthia Schuck-Paim and Alex Kacelnik. Rationality in risk-sensitive foraging choices by starlings. *Animal Behaviour*, 64(6):869–879, 2002.
- [7] V Arcis and D Desor. Influence of environment structure and food availability on the foraging behaviour of the laboratory rat. *Behavioural processes*, 60(3):191–198, 2003.
- [8] Evan C Carter and A David Redish. Rats value time differently on equivalent foraging and delay-discounting tasks. *Journal of Experimental Psychology: General*, 145(9):1093, 2016.
- [9] Lisa M Rose. Sex differences in diet and foraging behavior in white-faced capuchins (*cebus capucinus*). *International Journal of Primatology*, 15(1):95–114, 1994.
- [10] PA Garber. Foraging strategies among living primates. *Annual review of Anthropology*, 16(1):339–364, 1987.
- [11] Thomas T Hills. Animal foraging and the evolution of goal-directed cognition. *Cognitive science*, 30(1):3–41, 2006.
- [12] David W Stephens, Joel S Brown, and Ronald C Ydenberg. *Foraging: behavior and ecology*. University of Chicago Press, 2008.
- [13] Adam J Calhoun and Benjamin Y Hayden. The foraging brain. *Current Opinion in Behavioral Sciences*, 5:24–31, 2015.
- [14] Nigel E Raine, Thomas C Ings, Anna Dornhaus, Nehal Saleh, and Lars Chittka. Adaptation, genetic drift, pleiotropy, and history in the evolution of bee foraging behavior. *Advances in the Study of Behavior*, 36:305–354, 2006.
- [15] Marla B Sokolowski, H Sofia Pereira, and Kimberley Hughes. Evolution of foraging behavior in drosophila by density-dependent selection. *Proceedings of the National Academy of Sciences*, 94(14):7373–7377, 1997.
- [16] BJ King. Extractive foraging and the evolution of primate intelligence. *Human evolution*, 1(4):361, 1986.
- [17] Dean Mobbs, Pete C Trimmer, Daniel T Blumstein, and Peter Dayan. Foraging for foundations in decision neuroscience: Insights from ethology. *Nature Reviews Neuroscience*, 19(7):419, 2018.
- [18] Benjamin Y Hayden and Mark E Walton. Neuroscience of foraging. *Frontiers in neuroscience*, 8:81, 2014.
- [19] Sam Hall-McMaster and Fabrice Luyckx. Revisiting foraging approaches in neuroscience. *Cognitive, Affective, & Behavioral Neuroscience*, 19(2):225–230, 2019.
- [20] Jeffrey K Waage. Foraging for patchily-distributed hosts by the parasitoid, *nemeritis canescens*. *The Journal of Animal Ecology*, pages 353–371, 1979.
- [21] John McNamara. Optimal patch use in a stochastic environment. *Theoretical Population Biology*, 21(2):269–288, Apr 1982.
- [22] Eric L Charnov. Optimal foraging, the marginal value theorem. *Theoretical population biology*, 9(2):129–136, 1976.
- [23] Roger L Mellgren. Foraging in a simulated natural environment: There’s a rat loose in the lab. *Journal of the Experimental Analysis of Behavior*, 38(1):93–100, 1982.
- [24] Innes Cuthill and Alejandro Kacelnik. Central place foraging: a reappraisal of the ‘loading effect’. *Animal Behaviour*, 40(6):1087–1101, 1990.
- [25] H. Rita and E. Ranta. Stochastic patch exploitation model. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1393):309–315, Feb 1998.
- [26] Gerard Driessen and Carlos Bernstein. Patch departure mechanisms and optimal host exploitation in an insect parasitoid. *Journal of Animal Ecology*, 68(3):445–459, 1999.
- [27] Jane F Ward, Roger M Austin, and David W Macdonald. A simulation model of foraging behaviour and the effect of predation risk. *Journal of Animal Ecology*, 69(1):16–30, 2000.
- [28] Sigrunn Eliassen, Christian Jørgensen, Marc Mangel, and Jarl Giske. Quantifying the adaptive value of learning in foraging behavior. *The American Naturalist*, 174(4):478–489, Oct 2009.
- [29] Dale E Taneyhill. Patch departure behavior of bumble bees: rules and mechanisms. *Psyche: A Journal of Entomology*, 2010, 2010.
- [30] Feng Zhang and Cang Hui. Recent experience-driven behaviour optimizes foraging. *Animal behaviour*, 88:13–19, 2014.
- [31] Jacob D Davidson and Ahmed El Hady. Foraging as an evidence accumulation process. *PLoS computational biology*, 15(7):e1007060, 2019.
- [32] Jay M Biernaskie, Steven C Walker, and Robert J Gegear. Bumblebees learn to forage like bayesians. *The American Naturalist*, 174(3):413–423, 2009.
- [33] Richard F. Green. A simpler, more general method of finding the optimal foraging strategy for bayesian birds. *Oikos*, 112(2):274–284, 2006.
- [34] John M McNamara, Richard F Green, and Ola Olsson. Bayes’ theorem and its applications in animal behaviour. *Oikos*, 112(2):243–251, 2006.
- [35] Ola Olsson and Noël MA Holmgren. The survival-rate-maximizing policy for bayesian foragers: wait for good news. *Behavioral Ecology*, 9(4):345–353, 1998.
- [36] John McNamara and Alasdair Houston. The application of statistical decision theory to animal behaviour. *Journal of Theoretical Biology*, 85(4):673–690, Aug 1980.
- [37] Ola Olsson. Bayesian foraging with only two patch types. *Oikos*, 112(2):285–297, 2006.
- [38] Jean-Sébastien Pierre and Richard F Green. A bayesian approach to optimal foraging in parasitoids. *Behavioral ecology of insect parasitoids*. Blackwell, Oxford, pages 357–383, 2008.
- [39] Ola Olsson and Noël MA Holmgren. Gaining ecological information about bayesian foragers through their behaviour. i. models with predictions. *Oikos*, pages 251–263, 1999.
- [40] Jan A Van Gils. State-dependent bayesian foraging on spatially autocorrelated food distributions. *Oikos*, 119(2):237–244, 2010.
- [41] Miguel A Rodriguez-Girones and Rodrigo A Vasquez. Density-dependent patch exploitation and acquisition of environmental information. *Theoretical Population Biology*, 52(1):32–42, 1997.
- [42] Aaron M Ellison. Bayesian inference in ecology. *Ecology*

- letters, 7(6):509–520, 2004.
- [43] Juan C Alonso, Javier A Alonso, Luis M Bautista, and Rodrigo Muñoz-Pulido. Patch use in cranes: a field test of optimal foraging predictions. *Animal Behaviour*, 49(5):1367–1379, 1995.
- [44] Richard F Green. Bayesian birds: a simple example of oaten’s stochastic model of optimal foraging. *Theoretical Population Biology*, 18(2):244–256, 1980.
- [45] OLA Olsson, Ulf Wiklander, Noël MA Holmgren, and Sven G Nilsson. Gaining ecological information about bayesian foragers through their behaviour. ii. a field test with woodpeckers. *Oikos*, pages 264–276, 1999.
- [46] Thomas J. Valone. Are animals capable of bayesian updating? an empirical review. *Oikos*, 112(2):252–259, 2006.
- [47] Peter R Killeen, Gina-Marie Palombo, Lawrence R Gotlob, and Jon Beam. Bayesian analysis of foraging by pigeons (columba livia). *Journal of Experimental Psychology: Animal Behavior Processes*, 22(4):480, 1996.
- [48] Peter Nonacs and Joanne L Soriano. Patch sampling behaviour and future foraging expectations in argentine ants, *linepithema humile*. *Animal behaviour*, 55(3):519–527, 1998.
- [49] Thomas J Valone and Joel S Brown. Measuring patch assessment abilities of desert granivores. *Ecology*, 70(6):1800–1810, 1989.
- [50] Joshua I Gold and Michael N Shadlen. The neural basis of decision making. *Annual review of neuroscience*, 30, 2007.
- [51] Bingni W Brunton, Matthew M Botvinick, and Carlos D Brody. Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128):95–98, 2013.
- [52] Jan Drugowitsch, Rubén Moreno-Bote, Anne K Churchland, Michael N Shadlen, and Alexandre Pouget. The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11):3612–3628, 2012.
- [53] Jan Drugowitsch, Rubén Moreno-Bote, and Alexandre Pouget. Optimal decision-making with time-varying evidence reliability. In *Advances in neural information processing systems*, pages 748–756, 2014.
- [54] Crispin W Gardiner. *Handbook of stochastic methods*, volume 3. Springer, 2004.
- [55] Rafal Bogacz, Peter T Hu, Philip J Holmes, and Jonathan D Cohen. Do humans produce the speed–accuracy trade-off that maximizes reward rate? *The Quarterly Journal of Experimental Psychology*, 63(5):863–891, 2010.
- [56] Christopher M Glaze, Alexandre LS Filipowicz, Joseph W Kable, Vijay Balasubramanian, and Joshua I Gold. A bias–variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour*, 2(3):213–224, 2018.
- [57] Innes C. Cuthill, Alejandro Kacelnik, John R. Krebs, Patsy Haccou, and Yoh Iwasa. Starlings exploiting decisions: the effect of recent experience on foraging decisions. *Animal Behaviour*, 40(4):625–640, October 1990.
- [58] Yannick Outreman, Anne Le Ralec, Eric Wajnberg, and Jean-Sébastien Pierre. Effects of within- and among-patch experiences on the patch-leaving decision rules in an insect parasitoid. *Behavioral Ecology and Sociobiology*, 58(2):208–217, June 2005.
- [59] Steven L Scott. A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, 26(6):639–658, 2010.
- [60] Jeffrey S. Banks and Rangarajan K. Sundaram. Switching costs and the Gittins index. *Econometrica*, 62(3):687–694, 1994.
- [61] Alan C Kamil and Sonja I Yoerg. Learning and foraging behavior. In *Ontogeny*, pages 325–364. Springer, 1982.
- [62] Klaus Regelmann. Learning to forage in a variable environment. *Journal of theoretical biology*, 120(3):321–329, 1986.
- [63] Sara M Constantino and Nathaniel D Daw. Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience*, 15(4):837–853, 2015.
- [64] Jan A Van Gils, Ingrid W Schenk, Oscar Bos, and Theunis Piersma. Incompletely informed shorebirds that face a digestive constraint maximize net energy gain when exploiting patches. *The American Naturalist*, 161(5):777–793, 2003.
- [65] Raymond HG Klaassen, Bart A Nolet, Jan A van Gils, and Silke Bauer. Optimal movement between patches under incomplete information about the spatial distribution of food items. *Theoretical Population Biology*, 70(4):452–463, 2006.
- [66] Jessica J Hughes, David Ward, and Michael R Perrin. Effects of substrate on foraging decisions by a namib desert gerbil. *Journal of Mammalogy*, 76(2):638–645, 1995.
- [67] Thomas J Valone. Bayesian and prescient assessment: foraging with pre-harvest information. *Animal Behaviour*, 41(4):569–577, 1991.
- [68] Elizabeth A. Marschall, Peter L. Chesson, and Roy A. Stein. Foraging in a patchy environment: prey-encounter rate and residence time distributions. *Animal Behaviour*, 37:444–454, March 1989.
- [69] Alex Kacelnik and Melissa Bateson. Risky theories—the effects of variance on foraging decisions. *American Zoologist*, 36(4):402–434, 1996.
- [70] Vaibhav Srivastava, Paul Reverdy, and Naomi E. Leonard. On optimal foraging and multi-armed bandits. In *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 494–499, Monticello, IL, October 2013. IEEE.
- [71] Sudipto Guha and Kamesh Munagala. Multi-armed Bandits with Metric Switching Costs. In *ICALP*, 2009.
- [72] M. Asawa and D. Teneketzis. Multi-armed bandits with switching penalties. *IEEE Transactions on Automatic Control*, 41(3):328–348, March 1996.
- [73] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic Multi-Armed-Bandit Problem with Non-stationary Rewards. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 199–207. Curran Associates, Inc., 2014.
- [74] D.E. Koulouriotis and A. Xanthopoulos. Reinforcement learning and evolutionary algorithms for non-stationary multi-armed bandit problems. *Applied Mathematics and Computation*, 196(2):913–922, March 2008.
- [75] Lai Wei and Vaibhav Srivatsva. On Abruptly-Changing and Slowly-Varying Multiarmed Bandit Problems. In *2018 Annual American Control Conference (ACC)*, pages 6291–6296, June 2018. ISSN: 2378-5861.
- [76] Cédric Hartland, Sylvain Gelly, Nicolas Baskiotis, Olivier Teytaud, and Michèle Sebag. Multi-armed Ban-

- dit, Dynamic Environments and Meta-Bandits. In *nIPS-2006 Workshop, Online Trading Between Exploration and Exploitation*, volume hal-00113668, Whistler, Canada, 2006.
- [77] JG Ollason. Learning to forage in a regenerating patchy environment: can it fail to be optimal? *Theoretical Population Biology*, 31(1):13–32, 1987.
- [78] Alexandra T Pietrewicz and Jerry B Richards. *Learning to forage: an ecological perspective*. Lawrence Erlbaum Associates, 1985.
- [79] Michael Davison and Dianne McCarthy. *The matching law: A research review*. Routledge, 2016.
- [80] Nils Kolling and Thomas Akam. (reinforcement?) learning to forage optimally. *Current Opinion in Neurobiology*, 46:162–169, Oct 2017.
- [81] Athena Akrami, Charles D Kopec, Mathew E Diamond, and Carlos D Brody. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, 554(7692):368–372, 2018.
- [82] Ofri Raviv, Merav Ahissar, and Yonatan Loewenstein. How recent history affects perception: the normative approach and its heuristic approximation. *PLoS computational biology*, 8(10), 2012.
- [83] Paymon Ashourian and Yonatan Loewenstein. Bayesian inference underlies the contraction bias in delayed comparison tasks. *PloS one*, 6(5), 2011.
- [84] Masayuki Koganezawa, Hiroaki Hara, Yoshinori Hayakawa, and Ichiro Shimada. Memory effects on scale-free dynamics in foraging drosophila. *Journal of theoretical biology*, 260(3):353–358, 2009.
- [85] Claude Bélisle and James Cresswell. The effects of a limited memory capacity on foraging behavior. *Theoretical Population Biology*, 52(1):78–90, 1997.
- [86] Alan C. Kamil and Kevin C. Clements. Learning, memory, and foraging behavior. In *Contemporary issues in comparative psychology*, pages 7–30. Sinauer Associates, Sunderland, MA, US, 1990.
- [87] Melissa Bateson and Alex Kacelnik. Accuracy of memory for amount in the foraging starling, *sturnus vulgaris*. *Animal Behaviour*, 50(2):431–443, 1995.
- [88] Chloe Bracis, Eliezer Gurarie, and R Bram Van Moorter. Memory effects on movement behavior in animal foraging. *PloS one*, 10(8), 2015.
- [89] Alexander Huk, Kathryn Bonnen, and Biyu J He. Beyond trial-based paradigms: Continuous behavior, ongoing neural activity, and natural stimuli. *Journal of Neuroscience*, 38(35):7551–7558, 2018.
- [90] David L Barack, Steve WC Chang, and Michael L Platt. Posterior cingulate neurons dynamically signal decisions to disengage during foraging. *Neuron*, 96(2):339–347, 2017.
- [91] Benjamin Y Hayden, John M Pearson, and Michael L Platt. Neuronal basis of sequential foraging decisions in a patchy environment. *Nature neuroscience*, 14(7):933, 2011.
- [92] Eran Lottem, Dhruba Banerjee, Pietro Vertechi, Dario Sarra, Matthijs oude Lohuis, and Zachary F Mainen. Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nature communications*, 9(1):1–12, 2018.
- [93] Pietro Vertechi, Eran Lottem, Dario Sarra, Beatriz Godinho, Isaac Treves, Tiago Quendera, Matthijs Nicolai Oude Lohuis, and Zachary F Mainen. Inference-based decisions in a hidden state foraging task: Differential contributions of prefrontal cortical areas. *Neuron*, 2020.
- [94] David W Stephens. Decision ecology: foraging and the ecology of animal decision making. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4):475–484, 2008.
- [95] Andrew M Wikenheiser, David W Stephens, and A David Redish. Subjective costs drive overly patient foraging strategies in rats on an intertemporal foraging task. *Proceedings of the National Academy of Sciences*, page 201220738, 2013.
- [96] Eleanor Batty, Matthew Whiteway, Shreya Saxena, Dan Biderman, Taiga Abe, Simon Musall, Winthrop Gillis, Jeffrey Markowitz, Anne Churchland, John P Cunningham, Sandeep R Datta, Scott Linderman, and Liam Paninski. BehaveNet: nonlinear embedding and Bayesian neural decoding of behavioral videos. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d\textquotesingle Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 15706–15717, 2019.
- [97] Nils Kolling, Timothy EJ Behrens, Rogier B Mars, and Matthew FS Rushworth. Neural mechanisms of foraging. *Science*, 336(6077):95–98, 2012.
- [98] Amitai Shenhav, Mark A Straccia, Jonathan D Cohen, and Matthew M Botvinick. Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature neuroscience*, 17(9):1249, 2014.
- [99] Amitai Shenhav, Matthew M Botvinick, and Jonathan D Cohen. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, 79(2):217–240, 2013.
- [100] Harris S Kaplan and Manuel Zimmer. Brain-wide representations of ongoing behavior: a universal principle? *Current Opinion in Neurobiology*, 64:60–69, 2020.
- [101] David W Roubik. Comparative foraging behavior of apis mellifera and trigona corvina (hymenoptera: Apidae) on baltimora recta (compositae). *Revista de Biologia Tropical*, 29(2):177–183, 1981.
- [102] Frank J Bonaccorso, John R Winkelmann, Danny Shin, Caroline I Agrawal, Nadia Aslami, Caitlin Bonney, Andrea Hsu, Phoebe E Jekielek, Allison K Knox, Stephen J Kopach, et al. Evidence for exploitative competition: comparative foraging behavior and roosting ecology of short-tailed fruit bats (phyllostomidae). *Biotropica*, 39(2):249–256, 2007.
- [103] Eric Wajnberg, Pierre-Alexis Gonsard, Elisabeth Tabone, Christine Curty, Nathalie Lezcano, and Stefano Colazza. A comparative analysis of patch-leaving decision rules in a parasitoid family. *Journal of Animal Ecology*, pages 618–626, 2003.
- [104] Vaibhav Srivastava, Paul Reverdy, and Naomi Ehrich Leonard. Correlated multiarmed bandit problem: Bayesian algorithms and regret analysis. *arXiv preprint arXiv:1507.01160*, 2015.
- [105] Ran Nathan. An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences*, 105(49):19050–19051, 2008.
- [106] Brian J Jackson, Gusti Lulu Fatima, Sujean Oh, and David Henry Gire. Many paths to the same goal: meta-heuristic operation of brains during natural behavior. *bioRxiv*, page 697607, 2019.
- [107] Paul Cisek, Geneviève Aude Puskas, and Stephany El-Murr. Decisions in changing conditions: the urgency-gating model. *Journal of Neuroscience*, 29(37):11560–

- 11571, 2009.
- [108] Gaurav Malhotra, David S Leslie, Casimir JH Ludwig, and Rafal Bogacz. Time-varying decision boundaries: insights from optimality analysis. *Psychonomic bulletin & review*, 25(3):971–996, 2018.
- [109] Massimo Vergassola, Emmanuel Villermaux, and Boris I Shraiman. ‘infotaxis’ as a strategy for searching without gradients. *Nature*, 445(7126):406–409, 2007.
- [110] Beat Naef-Daenzer. Patch time allocation and patch sampling by foraging great and blue tits. *Animal behaviour*, 59(5):989–999, 2000.
- [111] David William Stephens. *Stochasticity in foraging theory: risk and information*. PhD thesis, University of Oxford, 1982.
- [112] SM Dow and SEG Lea. Sampling of schedule parameters by pigeons: tests of optimizing theory. *Animal behaviour*, 35(1):102–114, 1987.
- [113] Donald L Kramer and Daniel M Weary. Exploration versus exploitation: a field study of time allocation to environmental tracking by foraging chipmunks. *Animal Behaviour*, 41(3):443–449, 1991.
- [114] Simon A Levin, Thomas M Powell, and John H Steele. *Patch dynamics*, volume 96. Springer Science & Business Media, 2012.
- [115] Daniel Grünbaum. The logic of ecological patchiness. *Interface Focus*, 2(2):150–155, 2012.
- [116] Charles H Greene. Patterns of prey selection: implications of predator foraging tactics. *The American Naturalist*, 128(6):824–839, 1986.
- [117] Colin W Clark and Marc Mangel. Foraging and flocking strategies: information in an uncertain environment. *The American Naturalist*, 123(5):626–641, 1984.
- [118] Luc-Alain Giraldeau and Thomas Caraco. *Social foraging theory*, volume 73. Princeton University Press, 2018.
- [119] DGC Harper. Competitive foraging in mallards: “ideal free” ducks. *Animal Behaviour*, 30(2):575–584, 1982.
- [120] Jennifer J Templeton and Luc-Alain Giraldeau. Vicarious sampling: the use of personal and public information by starlings foraging in a simple patchy environment. *Behavioral Ecology and Sociobiology*, 38(2):105–114, 1996.
- [121] Alfonso Pérez-Escudero and Gonzalo G. de Polavieja. Collective Animal Behavior from Bayesian Estimation and Probability Matching. *PLoS Computational Biology*, 7(11):e1002282, November 2011.