Stochastic models of exploration in patch foraging tasks

By

Shayan Shafquat

School of Psychology
University of Nottingham

With Prof. Mark Humphries

Sep 2024

Word count: 8000 thousand

Abstract

Recent advances in decision neuroscience have used foraging-based studies to explore decision-making processes. Building on this research, we applied a well-established patch-foraging framework to study human decision-making in foraging contexts. Our findings replicate earlier results, showing that human foragers deviate from optimal models like the Marginal Value Theorem (MVT). To address the limitations of deterministic MVT, we investigated stochastic action-selection algorithms, including epsilon-greedy, softmax, and mellowmax, assessing their ability to model foraging dynamics like the patch effect and environmental influences. We adjusted parameters to examine how these models balance exploration and exploitation to align with optimal patch-leaving times. Analysis of subject-specific data from Le Heron's study revealed that the softmax model required an additional bias term to account for individual differences, while the mellowmax model, with adaptive parameters, better captured diverse foraging behaviors. Lastly, we introduced methods incorporating uncertainty into reward decay rates, shedding light on the role of uncertainty in driving stochasticity in patch-leaving decisions.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Mark Humphries, for his unwavering support, guidance, and invaluable insights throughout the course of this research. I am also deeply thankful to Emma Scholey for her pivotal assistance in establishing the foundation for the Marginal Value Theorem (MVT), which was central to this work. The contributions of those mentioned have been instrumental in the successful completion of this thesis.

List of Figures

Figure **Page**

Table of Contents

Chapter 1

Introduction

Decision-making is a fundamental cognitive process shaping behaviors in both humans and animals, such as deciding when to switch tasks or seek new resources [1, 2]. A key challenge is balancing the exploitation of known resources with the exploration of uncertain, potentially more rewarding alternatives [3, 4]. This balance is often modeled through foraging theory, which examines decisions about whether to continue extracting diminishing resources from a current patch or explore new ones [5, 6]. These processes are relevant across disciplines, including psychology, economics, ecology, and neuroscience, reflecting broader cognitive strategies for decision-making under uncertainty [7, 8, 9, 10, 11].

The Marginal Value Theorem (MVT) is a key model for optimal foraging behavior, proposing that individuals should leave a resource patch once the rate of return drops below the environmental average [6]. However, the MVT assumes perfect knowledge and rational behavior, conditions often absent in real-world scenarios [5, 12]. Both humans and animals frequently deviate from this model, exhibiting behaviors like overharvesting, where they remain in a patch longer than predicted [13, 14]. This raises questions about the cognitive mechanisms driving such behavior and the role of imperfect information and uncertainty in decision-making [15, 16].

Contemporary research increasingly adopts stochastic decision-making models that incorporate randomness, offering a more accurate representation of real-world behavior by introducing probabilistic elements [17, 18, 19]. These models are useful for explaining phenomena like overharvesting, often driven by uncertainty about future rewards or incomplete information [20]. In foraging studies, stochastic models often use algorithms

like softmax and epsilon-greedy to balance exploration and exploitation [3, 4, 21].

Exploration strategies in uncertain environments are key to foraging decisions. Directed exploration involves a strategic search for new resources based on environmental cues, guided by knowledge or inference about potential rewards [22, 23, 24]. In contrast, random exploration is a more indiscriminate search, often driven by internal factors like curiosity or frustration [25, 26, 27]. Both strategies are essential for navigating environments with unpredictable resources [28, 29].

This work examines the role of stochastic decision-making in foraging, particularly overharvesting. By using probabilistic models like softmax and mellowmax, the study aims to capture the complexity of real-world decisions better than deterministic models like the MVT [20, 30, 31]. Drawing on data from Le Heron et al.'s patch-foraging experiment [15], the research evaluates these models' effectiveness in explaining patch-leaving decisions and explores how uncertainty shapes stochasticity in decision-making. This study seeks to deepen our understanding of the cognitive mechanisms underlying decision-making in uncertain environments [11, 32].

Chapter 2

Methods

This chapter will detail the design and rationale of the Leheron's patch foraging task [15], examine the stochastic models used, explore their behavioural implications, outline methodologies for model fitting and comparison, and discuss methods for representing uncertainty.

2.1 Foraging Theory and Experimental Design in Decision-Making

This section outlines the design of Le Heron's patch foraging task, a simulation-based experiment developed to investigate decision-making in environments with depleting resources. The task is based on Optimal Foraging Theory (OFT) and the MVT, which predict the optimal time an agent should remain in a resource patch before leaving [6, 33]. Additionally, we describe the computational methods used to determine optimal leaving times, accounting for patch travel times and the complexities of realistic foraging environments.

2.1.1 Le Heron's Patch Foraging Task: Experimental Design and Rationale

The Le Heron's task is rooted in foraging theory and behavioral ecology [5, 6], simulating decision-making in depleting resource environments to test OFT and MVT predictions. By generating empirical data, it allows comparison with theoretical models, offering a platform to explore adaptive foraging strategies [9, 13].

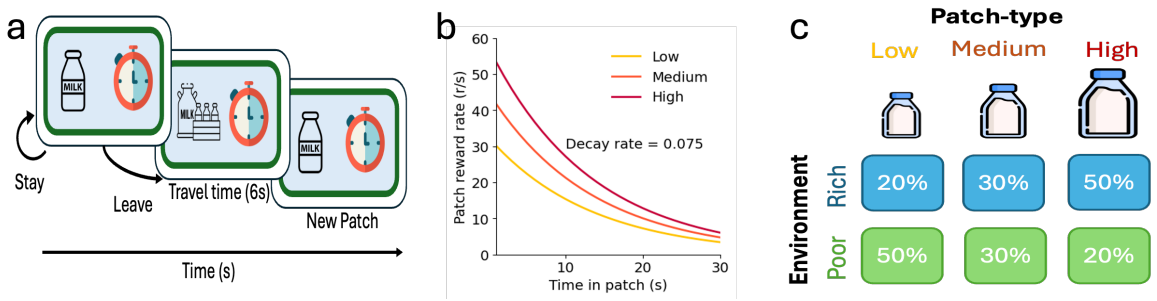

Figure 2.1: **Le Heron's patch foraging task setup.** The figure shows Low, Medium, and High patches (left), reward decay for different initial reward rates (center), and patch distribution in Rich and Poor environments (right).

Human participants engage with virtual patches offering diminishing rewards, analogous to real-world foraging where resources become scarce over time [5]. In this task, "milk" symbolizes the reward, which participants collect by deciding how long to stay in a patch before moving on. Rewards decrease exponentially (Figure 2.1.b), modeled as:

$$r(t) = r_0 \exp(-\alpha t) \tag{2.1}$$

where r_0 is the initial reward rate and α is the decay constant decay rate. Notably, α remains uniform across all patch types, allowing us to isolate the effects of r_0 on particpant's leave or stay decisions.

The task categorizes patches as Low, Medium, or High based on initial reward rate r_0. High patches offer the highest r_0, providing greater immediate returns, while Low patches offer the least. Medium patches offer a moderate initial reward rate. This design enables the analysis of how different initial conditions affect leaving decisions [9].

Additionally, the experiment includes Rich and Poor environments, which differ in patch distribution. Rich environments contain more High reward patches, while Poor environments feature more Low reward patches (Figure 2.1.c), simulating different levels of resource abundance and enabling investigation into how environmental richness influences foraging behavior [34, 35].

2.1.2 Computational Approach to Optimal Leaving Times

This computational method, based on personal communication, was implemented in Python and extended to include both exponential and linear reward decay models. While

the MVT provides a theoretical basis for determining optimal leaving times in foraging tasks [6], decision-making in the Le Heron's task necessitates a computational approach due to the complexity of estimating the average reward rate across diverse patch types and environments. This complexity arises from accounting for both within-patch reward dynamics and between-patch travel times [9]. The goal is to compute optimal leaving times for each patch type, considering reward decay within patches and travel times between them.

Reward Rate for Individual Patches

The decision to leave a patch depends on the reward rate $\mathbf{RR}_i(t)$, defined as the ratio of accumulated gain to total time spent in the patch, including travel time [5]:

$$\mathbf{RR}_i(t) = \frac{\mathbf{Gain}_i(t)}{t_{\text{travel}} + t}, \quad i = 1, 2, 3$$

where $\mathbf{Gain}_i(t)$ represents accumulated gain at time t, and t_{travel} is the travel time, set to 6 unit. For patches with exponentially decaying rewards, the gain $\mathbf{Gain}_i(t)$ is:

$$\mathbf{Gain}_i(t) = \frac{r_{0i}}{\alpha}\left(1 - \exp(-\alpha t)\right)$$

where r_{0i} is the initial reward of patch P_i, and α is the decay constant.

Multi-Patch Environment and Overall Reward Rate

In multi-patch environments, the overall reward rate depends on time spent in each patch type and their proportions. Let $\mathbf{p} \in \mathbb{R}^3$ represent the proportions of Low, Medium, and High resource patches:

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}, \quad p_1 + p_2 + p_3 = 1$$

The multi-patch reward rate matrix $\mathbf{multiPatchRR} \in \mathbb{R}^{T\times 1\times 3}$, represents reward rates across all patch types and time steps. Here T is the number of time steps, and 3 corresponds to the patch types. Next, we construct the current patch reward rate matrix

curPatchRR $\in \mathbb{R}^{T\times T\times T\times 3}$, tiling **multiPatchRR** across the appropriate dimensions for Low, Medium, and High patches.

The overall reward rate matrix **OverallRR** $\in \mathbb{R}^{T\times T\times T\times 1}$ aggregates the reward rates across all combinations of time spent in each patch type and the travel time. Here, 1 corresponds to the size of travel time array, as the travel time is consistent across patches. The overall reward rate $\mathbf{OverallRR}(t_1, t_2, t_3)$ is computed as:

$$\begin{aligned}\mathbf{OverallRR}(t_1, t_2, t_3) = p_1 \cdot \mathbf{curPatchRR}_1(t_1) \\ + p_2 \cdot \mathbf{curPatchRR}_2(t_2) \\ + p_3 \cdot \mathbf{curPatchRR}_3(t_3)\end{aligned} \tag{2.2}$$

where t_1, t_2, t_3 represent the time spent in the Low, Medium, and High resource patches, respectively. The average reward rate is obtained by maximizing this matrix across all time combinations.

Optimal Leaving Times

The optimal leaving times t_1^*, t_2^*, t_3^* are chosen as:

$$(t_1^*, t_2^*, t_3^*) = \arg\max_{t_1,t_2,t_3} \mathbf{OverallRR}(t_1, t_2, t_3)$$

This process identifies the time steps t_1^*, t_2^*, t_3^* that maximize the overall reward rate, balancing resource depletion and travel time. This method aligns with the principles of OFT [33] and provides a computational implementation of the MVT for multi-patch environments.

Adaptation to Linear Decay

In the linear decay model, resources deplete at a constant rate. The reward function is:

$$r(t) = r_0 - \alpha t$$

and the accumulated gain is:

$$\mathbf{Gain}_i(t) = r_0 t - \frac{1}{2}\alpha t^2$$

These equations can be substituted into the reward rate matrix $\mathbf{RR}_i(t)$, and the same optimization procedure is used to determine the optimal leaving times. Although this approach simplifies real-world foraging scenarios, it provides a solid computational framework for determining optimal foraging strategies [30, 21].

2.2 Stochastic Decision-Making Models

Foraging behavior in both humans and animals often deviates from deterministic models like the MVT due to inherent randomness in decision-making [5]. While MVT offers a clear rule for optimal patch-leaving times, it fails to capture the probabilistic nature of real-world behavior [6]. Stochastic models introduce variability into decision-making, providing a more realistic framework [9]. This section explores three key stochastic models—epsilon-greedy, softmax, and mellowmax—each addressing uncertainty in distinct ways.

2.2.1 Epsilon-greedy Strategy

The epsilon-greedy model introduces randomness by selecting the best-known option most of the time while occasionally exploring alternatives with a small probability ϵ [36, 37]. At each timestep t , the agent either remains in the current patch with probability $1 - \epsilon$ or explores a new patch with probability ϵ.

The parameter ϵ , ranging from 0 to 1, controls the exploration likelihood, while t represents the timestep. The leaving probability, $\pi_{\text{epsilon-greedy}}(\text{leave} \mid t)$, is defined as:

$$\pi_{\text{epsilon-greedy}}(\text{leave} \mid t) = \epsilon$$

Although this strategy encourages exploration, its simplicity limits adaptability in environments with changing reward dynamics, making it less effective in such scenarios [38]. Nonetheless, it serves as a useful benchmark for evaluating more complex models [39].

2.2.2 Softmax Strategy

The softmax model provides a more dynamic approach by making the leaving probability dependent on the current reward rate. It is governed by the inverse temperature parameter β , which controls the trade-off between exploration and exploitation [4]. Higher values of β increase exploitation, while lower values promote exploration [3]:

$$\pi_{\text{softmax}}(\text{leave} \mid t) = \frac{1}{1 + \exp(\beta r(t))}$$

Modified softmax: Bias term

An intercept term c can be added to adjust the decision boundary, allowing the model to account for factors beyond the immediate reward rate:

$$\pi_{\text{softmax}}(\text{leave} \mid t) = \frac{1}{1 + \exp(c + \beta r(t))}$$

This adjustment makes the softmax model more flexible in representing decision-making in variable conditions, including individual differences across subjects [30].

2.2.3 Mellowmax Strategy

The mellowmax strategy introduces a novel way to balance exploration and exploitation in decision-making by smoothing sensitivity to extreme reward values through a logarithmic transformation [31].

Mellowmax Operator

In the foraging context, the agent's decision to stay or leave uses the binary-action value function Q(t, a) , defined as:

$$Q(t, a) = \begin{cases} r(t) & \text{if } a = \text{stay} \\ 0 & \text{if } a = \text{leave} \end{cases} \tag{2.3}$$

where $r(t)$ is the reward at timestep t.

The mellowmax operator adapted to this stay-or-leave decision by incorporating the value functions defined in (2.3):

$$\mathrm{mm}_\omega(Q(t,\cdot)) = \frac{\log\left(1 + e^{\omega r(t)}\right)}{\omega} \tag{2.4}$$

where ω controls the smoothness of the operator.

Optimization Step

To adapt to the environment, the agent's decision-making is optimized at each timestep by adjusting the parameter β , balancing the decision to stay or leave. This is achieved through numerical methods like Brent's optimization [40], by solving:

$$\sum_{a\in\{\text{stay,leave}\}} e^{\beta(Q(t,a)-\mathrm{mm}_\omega(Q(t,\cdot)))}\left(Q(t,a) - \mathrm{mm}_\omega(Q(t,\cdot))\right) = 0 \tag{2.5}$$

Once β is optimized, the agent's probability of leaving at time t is given by:

$$\pi_{\text{mellowmax}}(\text{leave} \mid t) = \frac{1}{1 + e^{\hat{\beta} r(t)}} \tag{2.6}$$

where $\hat{\beta}$ is the optimized parameter, allowing the agent to adjust its decision based on the reward [31].

Modified Mellowmax Operator

A modified version of the mellowmax operator introduces a constant c, enhancing flexibility in decision-making [41]. The modified operator is given by:

$$\mathrm{mm}_\omega(Q(t,\cdot)) = \frac{\log\left(1 + e^{\omega r(t)} + c\right)}{\omega} \tag{2.7}$$

This modification provides more adaptability in balancing exploration and exploitation, particularly in environments with varying reward dynamics or when accounting for individual differences [4].

2.3 Behavioural Implications of Stochastic Models

Understanding the behavioural implications of stochastic models is crucial for predicting decision-making in patch-foraging tasks. Each model affects the timing of leaving decisions differently. This section explores how decision variability is influenced by the model used, with insights adapted from personal communications during this study.

2.3.1 Behaviour Under Epsilon-Greedy Strategy

The epsilon-greedy model results in a geometric distribution of leaving times, where the expected leaving time is:

$$E(\text{leave}) = \frac{1}{\epsilon} \tag{2.8}$$

The variance in leaving time is:

$$\text{VAR}(\text{leave}) = \frac{1-\epsilon}{\epsilon^2} \tag{2.9}$$

Since the model does not adapt to changes in reward dynamics, it serves as a benchmark rather than a comprehensive solution.

2.3.2 Behaviour Under Softmax and Mellowmax Strategies

In contrast, the softmax model is more adaptable to dynamic environments [37]. The probability of leaving a patch is a function of the current reward rate, modulated by β and c:

$$p(\text{leave}|t) = \frac{1}{1+\exp(c+\beta r_0 \exp(-\alpha t))} \tag{2.10}$$

The expected leaving time $E(\text{leave})$ and variance $\text{VAR}(\text{leave})$ in the softmax model are computed using the probability $p(\text{leave} = t)$, that the agent leaves at time t :

$$p(\text{leave} = t) = p(\text{leave}|t) \prod_{i=1}^{t-1} (1 - p(\text{leave}|i)) \tag{2.11}$$

where $p(\text{leave}|t)$ is the probability of leaving at time t, and the product accounts for the probability of not leaving at earlier time steps.

The expected leaving time $E(\text{leave})$ and variance VAR(leave) are given as:

$$E(\text{leave}) = \sum_{t=1}^{\infty} t \cdot p(\text{leave} = t) \tag{2.12}$$

$$\text{VAR}(\text{leave}) = \sum_{t=1}^{\infty} (t - E(\text{leave}))^2 \cdot p(\text{leave} = t) \tag{2.13}$$

These sums are computed numerically by truncating at a sufficiently large value of t, ensuring that $t_{\max}$ is much larger than $E(\text{leave})$.

The mellowmax model follows a similar calculation for expected leaving time and variance, given its probabilistic framework for decision-making similar to softmax as in (2.6).

2.4 Model Fitting and Comparison in Foraging Decision-Making

This section describes the methodology for fitting stochastic action selection models—specifically softmax and mellowmax operators—to empirical data from foraging experiments [15], followed by a comparative analysis of the models.

2.4.1 Analytical Approach to Model Fitting

An analytical approach was used to estimate the mean leaving times for different patch types as in (2.12), avoiding traditional log-likelihood maximization. Given the models' simplicity and the binary nature of the actions, this method enabled efficient parameter estimation.

For each subject, we fitted the parameters of the softmax and mellowmax models to their observed behavior. The predicted mean leaving times were calculated analytically, with parameters fitted individually for each participant.

Model Fitting Scenarios

We explored four model-fitting scenarios:

- **Case 1:** The parameters ($\beta/\omega/intercept$) were held constant across both environments (rich and poor) for each subject.
- **Case 2:** The parameters β/ω were fixed from Case 1, while the *intercept* varied between environments.
- **Case 3:** The *intercept* was fixed from Case 1, while parameters β/ω varied between environments.
- **Case 4:** The parameters β/ω varied freely between environments, while the *intercept* was fixed at zero.

2.4.1.1 Objective Function: Calculation and its Application to Scenarios

To evaluate model fit, we minimized the root mean square error (RMSE) between predicted and actual mean leaving times for each patch type, weighted by trial frequency:

$$\text{RMSE}^{(j)} = \sqrt{\frac{1}{N^{(j)}} \sum_{i=1}^{N^{(j)}} c_i \left(\mu_i^{\text{pred}}(\theta^{(j)}) - \mu_i^{\text{actual}} \right)^2}$$

where:

- $N^{(j)}$ is the number of patch types for subject j, which varies by case,
- c_i is the count of trials for patch type i,
- $\mu_i^{\text{pred}}(\theta^{(j)})$ is the predicted mean leaving time for patch type i based on parameters $\theta^{(j)}$,
- μ_i^{actual} is the empirical mean leaving time for patch type i,
- $\theta^{(j)}$ represents the fitted parameters (e.g., β, ω, *intercept*) for subject j.

For **Case 1**, the RMSE was computed across all trials, assuming constant parameters across environments. In **Cases 2-4**, the RMSE was computed separately for trials in rich and poor environments, allowing parameters to vary by environment:

$$\text{RMSE}^{(j)}_{\text{subject-env}} = \sqrt{\frac{1}{N^{(j)}_{\text{env}}} \sum_{i=1}^{N^{(j)}_{\text{env}}} c_i \left(\mu_i^{\text{pred}}(\theta^{(j)}_{\text{env}}) - \mu_i^{\text{actual}}\right)^2}$$

where $N^{(j)}_{\text{env}}$ is the number of patch types within each environment for subject j, and $\theta^{(j)}_{\text{env}}$ represents the parameters that are allowed to vary by environment.

Optimization Procedure

The Scipy package was used to perform optimization using the Nelder-Mead method, chosen for its robustness in handling non-linear models without parameter constraints [42]. The optimal parameter values for each case were then used to predict mean leaving times for rich and poor environments. This rigorous approach facilitated the evaluation of the models' ability to replicate individual decision-making in foraging contexts [5].

2.4.2 Bayesian Information Criterion for Model Comparison

The Bayesian Information Criterion (BIC) was used to compare models or parameterizations, balancing model fit with complexity [43]. BIC for each subject was calculated as follows:

$$\text{BIC}_j = k \cdot \ln(n_j) - 2 \cdot \ln(\hat{L}_j) \tag{2.14}$$

The likelihood $\hat{L}_j$ was derived from the residual sum of squares (RSS) for each subject. Assuming a Gaussian error model:

$$\ln(\hat{L}_j) = -\frac{n_j}{2}\left(\ln\left(2\pi \cdot \frac{\text{RSS}_j}{n_j}\right) + 1\right) \tag{2.15}$$

where:

$$\text{RSS}_j = \sum_{i=1}^{n_j} (y_{ij} - \hat{y}_{ij})^2$$

where y_{ij} is the observed value for patch sequence i of subject j, and $\hat{y}_{ij}$ is the the corresponding model prediction.

The log-likelihood from equation (2.15) was used in (2.14) to compute the BIC for each subject with $k = 1$ for Case 4 (no intercept) and $k = 2$ for the other cases. Individual BIC scores were summed across all subjects:

$$\text{Total BIC} = \sum_{j=1}^{S} \left(k \cdot \ln(n_j) - 2 \cdot \ln(\hat{L}_j) \right) \tag{2.16}$$

where S is the total number of subjects. Lower BIC scores indicated more favorable models that balance fit and complexity, a method validated in decision-making studies [44, 45].

2.5 Modeling Directed Exploration under Reward Uncertainty

Understanding how agents manage uncertainty in dynamic environments is crucial for elucidating decision-making processes. This section outlines the methods used to predict exploration strategy in foraging tasks where agents face uncertainty about the decay rate (θ) of rewards in resource patches.

2.5.1 Belief-Updating for Estimating Decay Rate

To model how agents handle uncertainty, various belief-updating mechanisms were implemented, including Conservative, Bayesian, Rescorla-Wagner and Incremental models. These mechanisms allow agents to refine their estimates of the decay rate (θ) over time, adjusting their beliefs based on data from each patch. Updates occur on both per-patch and per-environment bases when the agent decides to leave a patch. When encountering a patch type and environment for the first time, the agent samples θ from a Normal distribution where the mean is the actual decay rate.

Conservative Update: Preserving Initial Mean

The Conservative Update model allows the agent to maintain its initial belief about the mean decay rate while reducing uncertainty as more observations are made. The initial variance (σ^2_{prior}) is 0.015 updated according to:

$$\sigma^2_{\text{updated}} = \frac{\sigma^2_{\text{prior}}}{n}$$

where n is the number of observations. This approach suits environments where the agent has strong prior knowledge about θ and seeks to increase certainty without altering the mean belief [46].

Bayesian Update: Reducing Uncertainty Iteratively

In the Bayesian framework, the agent updates its belief about (θ) by integrating prior information with new observations. The belief is represented by a normal distribution, with the mean (μ) indicating the best estimate of θ and the variance (σ^2) reflecting uncertainty. The initial variance is 0.01 and the observation variance ($\sigma^2_{\text{observed}}$) is set to 0.00001, reflecting high confidence in observed data.

The posterior mean and variance after sampling a decay rate (θ_{sampled}) are calculated as:

$$\mu_{\text{posterior}} = \frac{\sigma^2 \theta_{\text{sampled}} + \sigma^2_{\text{observed}} \mu_{\text{prior}}}{\sigma^2 + \sigma^2_{\text{observed}}}$$
$$\sigma^2_{\text{posterior}} = \frac{\sigma^2 \sigma^2_{\text{observed}}}{\sigma^2 + \sigma^2_{\text{observed}}}$$

This mechanism is effective when the agent can gather substantial data from the environment [47].

Rescorla-Wagner Model: Prediction Error-Driven Learning

The Rescorla-Wagner model updates the agent's belief based on prediction error [48], defined as the difference between the observed reward (R_{obs}) and the expected reward (R_{exp}). The prediction error (δ) is calculated as:

$$\delta = R_{\text{obs}} - R_{\text{exp}}$$

The agent updates its belief about (θ) as follows:

$$\theta_{\text{new}} = \theta_{\text{old}} + \alpha \cdot \delta$$

Here, the learning rate (α) is set to 0.001, allowing slow adaptation. The initial variance is 0.01, and it remains constant throughout the process.

Incremental Update: Gradual Belief Adjustment

The Incremental update model uses a learning rule that promotes gradual adaptation of the agent's belief about (θ). The update rule is:

$$\theta_{\text{updated}} = \theta_{\text{prior}} + \alpha(\theta_{\text{sampled}} - \theta_{\text{prior}})$$

The learning rate (α) is 0.1, enabling faster adaptation than in the Rescorla-Wagner model. The initial variance is 0.015, and like the Rescorla-Wagner model, the variance remains unchanged throughout the learning.

2.5.2 Simulation and Decision-Making Process

The belief-updating mechanisms were integrated into a foraging simulation to model how agents handle uncertainty in resource environments. At the start of each trial, the agent sampled a decay rate (θ) from its belief distribution. As data were collected, the agent updated its beliefs about θ using one of the specified methods.

The agent's decision to leave a patch was governed by the MVT [6], which posits that an agent will leave a patch when the estimated reward rate ($r(t)$), derived from the decay rate θ falls below the average reward rate ($\bar{r}$) of the environment (calculated as in Section 2.1.2):

$$r(t) < \bar{r}$$

This decision-making framework allowed agenrs to adjust their behavior dynamically based on updated beliefs about the reward decay rates, balancing exploration and exploitation under uncertainty [3].

Chapter 3

Results

This chapter begins by examining the impact of patch rewards and environmental conditions on foraging dynamics, contrasting human behavior with predictions from Optimal Foraging Theory [5, 6]. We then assess the efficacy of various stochastic decision-making models, including the epsilon-greedy, softmax, and mellowmax approaches, in capturing the probabilistic aspects of foraging behavior [37]. Subsequently, through model fitting to empirical data, we explore individual differences in decision-making and how contextual factors, such as the environment, influence the balance between exploitation and exploration [3]. The chapter concludes with an in-depth analysis of variability in patch-leaving behavior, highlighting the role of uncertainty as a driving force behind stochastic decision-making processes, consistent with recent insights from the field of cognitive science [26, 49].

3.1 Patch-Leaving Behavior Across Patch Types

This section examines how foragers decide when to leave a patch under different patch types, reward decay models, and environmental conditions. We first applied a linear reward decay model to predict optimal leaving times, followed by an analysis using an exponential decay model based on Le Heron's patch-foraging experiment. Finally, these theoretical predictions, derived from optimal foraging theory (OFT), were compared with observed human behavior.

3.1.1 Dynamics of Patch Rewards and Environmental Influence

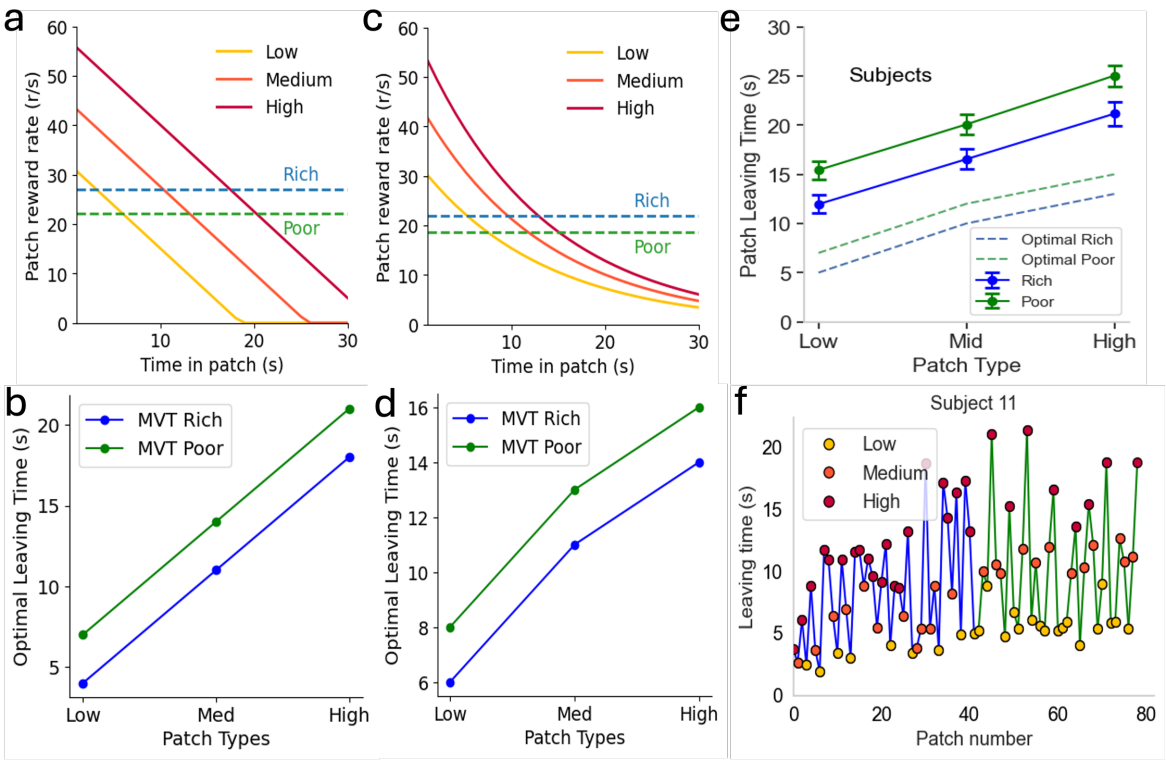

Figure 3.1: **Patch-leaving behavior across patch types and environmental conditions.** (a, c) Patch reward decay over time for Low, Medium, and High richness levels under linear and exponential decay models, respectively. Horizontal lines indicate average reward rates in rich and poor environments. (b, d) Optimal leaving times predicted using MVT for rich and poor environments across patch types under linear and exponential decay models, respectively. (e) Comparison of empirical patch-leaving times with MVT predictions. (f) Patch-leaving times for Subject 11, illustrating individual variability across patch richness levels.

Understanding patch reward dynamics is key to predicting foraging behavior. As patches are exploited, reward rates decline, influenced by initial patch richness (Low, Medium, High). Figures 3.1a and 3.1c show this decline for linear and exponential decay models, respectively. The Marginal Value Theorem (MVT) suggests foragers should leave a patch when its reward rate matches the environment's average rate [6]. Figures 3.1a and 3.1c show patches with higher yields maintain their reward rates above the environmental average for longer. This "patch effect" encourages longer patch stays for richer patches [9]. Figures 3.1b and 3.1d confirm this pattern across decay models and environments.

Environmental richness also significantly affects optimal leaving times. In richer

environments, characterized by higher average rewards, foragers leave patches earlier, as seen in Figures 3.1b and 3.1d. This "environment effect" highlights how foraging strategies adapt to varying environmental conditions [5, 34].

3.1.2 Human Behaviour Versus Optimal Foraging Predictions

The MVT assumes perfect knowledge of environmental reward rates, an unrealistic condition for human foragers [5]. Humans often deviate from MVT predictions, overharvesting and relying on heuristics rather than strict optimization [4, 30]. This is evident in Figure 3.1e, where human subjects in Le Heron's experiment (see Figure 2.1) stayed longer in patches than MVT predicted. Cognitive factors, including risk aversion and reward uncertainty may contribute to these deviations [32, 11].

Furthermore, the individual variability in patch-leaving times, as illustrated in Figure 3.1f, underscores the limitations of deterministic models like the MVT. Stochastic models, which account for randomness and individual differences, may better explain human decision-making [50, 51]. These findings emphasize the role of cognitive and environmental uncertainties in explaining deviations from optimal strategies [20, 52].

3.2 Stochastic Approaches in Patch-Leaving Decisions

The MVT provides a deterministic perspective on foraging behavior, predicting the optimal time to leave a patch based on the environment's average reward rate [6]. However, real-world decision-making often involves stochastic elements, such as incomplete knowledge of the environment or strategies that rely on randomness or exploration. To address this complexity, we evaluated stochastic models, including epsilon-greedy, softmax, and mellowmax-based action selection algorithms. These models provide insights into the exploration-exploitation trade-offs inherent in patch-leaving decisions [37].

3.2.1 Evaluation of Epsilon-Greedy Action Selection

The epsilon-greedy method, a classical reinforcement learning (RL) algorithm for balancing exploration and exploitation, was tested for its ability to model patch-leaving behavior [37]. As shown in Figure 3.2, we evaluated the algorithm using epsilon values of 0.08 and 0.1. These values govern the degree of exploration (random action selection) versus

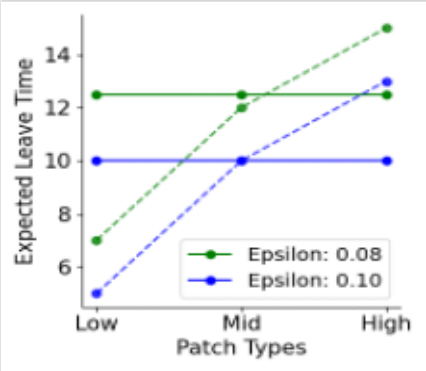

Figure 3.2: **Mean leave time across patch types for different epsilon values.** This plot compares the mean leave times for epsilon values of 0.08 and 0.1 against the optimal leave time predicted using MVT.

exploitation (greedy selection of the best-known option). While this method captures the environmental effect, it fails to account for differences for the patch effect. This limitation suggests that epsilon-greedy is not sufficiently robust for capturing the full spectrum of decision-making strategies in foraging tasks.

Though popular in RL due to its simplicity, the epsilon-greedy algorithm tends to oversimplify decision-making in complex environments with varying patch qualities. Our findings align with recent research advocating for more advanced models to capture human decision-making in uncertain environments [53].

3.2.2 Softmax Models: A Probabilistic Perspective

Softmax-based action selection offers a probabilistic framework where the probability of remaining in or leaving a patch depends on the rewards obtained at each time step. Two variations of this model were assessed: one incorporating an intercept (bias term, c) and one without. The first row of Figure 3.3 illustrates how the expected leaving times vary as a function of bias (c) and inverse temperature (β), based on methods described in Section 2.3.2. Figures 3.3a, 3.3c, and 3.3e demonstrate that an increase in either the bias term or inverse temperature leads to more exploitative behaviour.

The bottom row of Figure 3.3 compares the softmax model's expected leave times (solid lines) with the optimal MVT-predicted leave times (dashed lines). The expected leaving times were determined using calibrated values of β (β_{rich}, β_{poor}) and c (c_{rich}, c_{poor}) (shown by vertical dashed lines) demonstrating how miscalibrated parameters can result in overharvesting or underharvesting. Overall, the softmax model aligns more closely

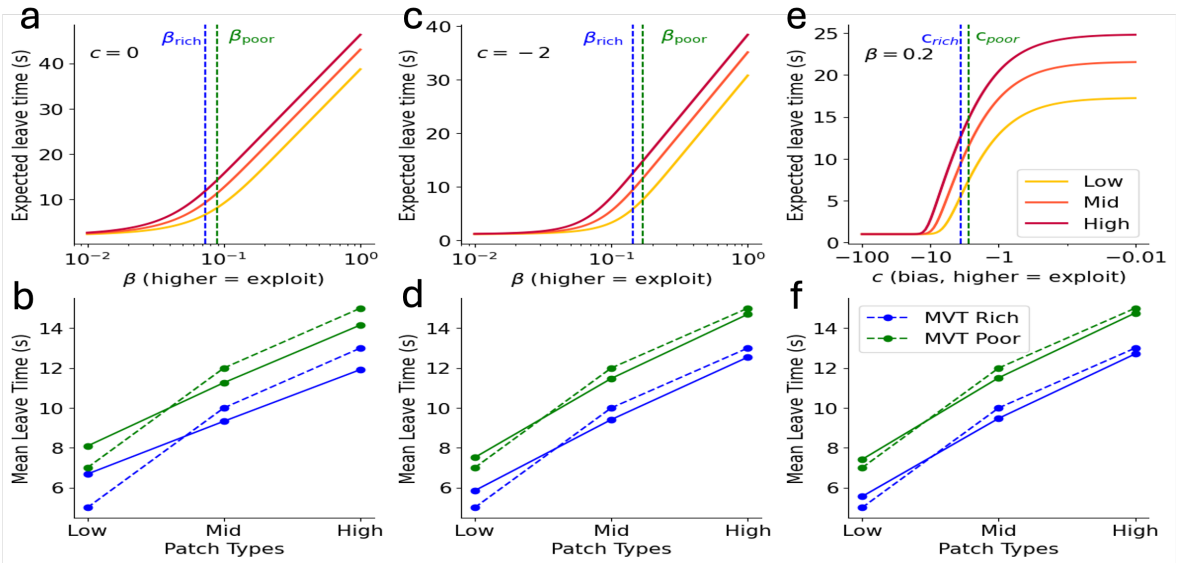

Figure 3.3: **Evaluation of Softmax-Based Models**. The top row (a, c, e) shows expected leave times as a function of bias (c) and inverse temperature β. The bottom row (b, d, f) compares softmax expected leave times (solid lines) with MVT predictions (dashed lines). Vertical dashed lines represent calibrated parameters (β_{rich}, β_{poor}) or (c_{rich}, c_{poor}), demonstrating that incorrect calibration can lead to overharvesting or underharvesting.

with MVT predictions than the epsilon-greedy approach.

These results support existing research, which highlights the adaptability of softmax models in environments with dynamic reward distributions [4, 54]. However, accurate parameter calibration remains essential to avoid deviations from optimal behavior.

3.2.3 Mellowmax-Based Action Selection: A Refined Strategy

The mellowmax-based algorithm, a refinement of the softmax approach, offers reduced sensitivity to reward differences, providing more stability in complex foraging tasks [31]. We evaluated mellowmax under similar conditions, with and without the inclusion of the bias term (c).

As shown in Figure 3.4, increasing ω produces more exploitative behavior, similar to the effect of increasing β in the softmax model. However, mellowmax reveals distinct patterns when the bias term (c) is varied while ω is held constant, especially in high-exploitation scenarios (Figure 3.4e).

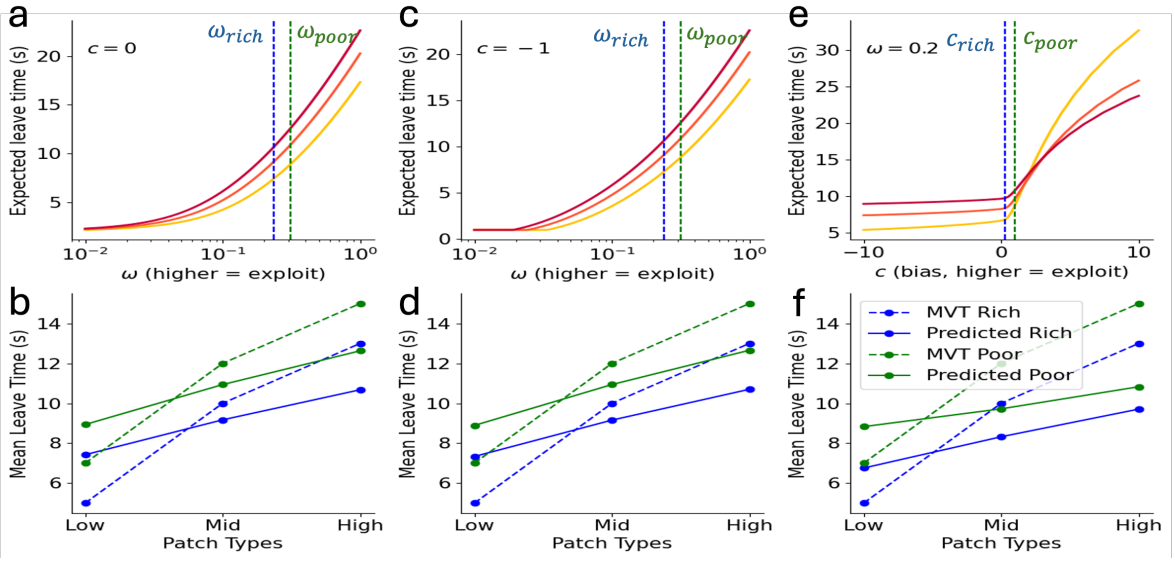

Figure 3.4: **Evaluation of Mellowmax-Based Models**. The top row (a, c, e) displays expected leave times as a function of ω and bias (c). The bottom row (b, d, f) compares mellowmax expected leave times (solid lines) with MVT predictions (dashed lines). Vertical dashed lines represent calibrated parameters (ω_{rich}, ω_{poor}) or (c_{rich}, c_{poor}), demonstrating that incorrect calibration can lead to overharvesting or underharvesting.

In the bottom row of Figure 3.4, the mellowmax model's expected leave times (solid lines) are compared with the MVT-predicted leave times (dashed lines). As with softmax, calibrated values of ω and c show how poor tuning leads to either overharvesting or underharvesting. Despite some misalignment with MVT predictions, the mellowmax model effectively captures both patch and environmental effects, particularly under certain parameter settings.

While mellowmax demonstrates weaker alignment with MVT predictions compared to softmax, both models perform well in dynamic environments by capturing key factors influencing forager decisions. These results are consistent with previous findings, which emphasize the value of probabilistic models in complex, stochastic environments [26].

3.3 Unveiling Individual Differences: Fitting Stochastic Models

Human decision-making is a complex process shaped by various factors, such as cognitive biases, personal differences, and environmental uncertainties [1]. While theoretical models

often make simplified assumptions, fitting these models to empirical data helps assess their ability to explain the underlying complexities. This approach also allows us to identify key model parameters that contribute to decision-making behaviour, offering insights into the cognitive processes involved [55].

In this study, we fitted parameters of two stochastic models (softmax and mellowmax) to the expected patch-leaving times of participants in Le Heron's patch-foraging experiment. Four different model-fitting scenarios were explored: fixing both β and c (intercept) in rich and poor environments, allowing one parameter to vary while keeping the other fixed, varying both β and c, and lastly, setting $c = 0$ for the two environments (see Methods 2.4.1). Through this process, we aimed to uncover how each model captured the balance between exploration and exploitation under different environmental conditions [3].

3.3.1 Capturing Individual Variability: Softmax vs. Mellowmax

We compared the softmax and mellowmax models to evaluate their efficacy in capturing individual variability in Le Heron's patch-foraging data. This analysis highlights the differences in how each model accounts for individual decision-making strategies.

To assess the models, we used the Bayesian Information Criterion (BIC), which balances accuracy with complexity, thereby preventing overfitting. The scenario with the lowest BIC sum across subjects indicated the most parsimonious model. As illustrated in Figures 3.5a and 3.5e, the best fit was achieved when the *intercept* was varied while either β (softmax) or ω (mellowmax) was held constant across different environments for each participant. In comparing the models, softmax slightly outperformed mellowmax, providing a marginally better fit overall. This is further supported by Figures 3.5b and 3.5f, where both models closely matched the empirical mean patch-leaving times across different environments.

More significant insights emerged when assessing individual variability (Figures 3.5c-d and 3.5g-h). The inclusion of an intercept in the softmax model considerably improved its ability to account for individual differences, as indicated by higher correlation coefficients r between predicted and observed leaving times. This improvement underscores the importance of considering individual biases and personal strategies in decision-making, consistent with research on foraging behaviour [56].

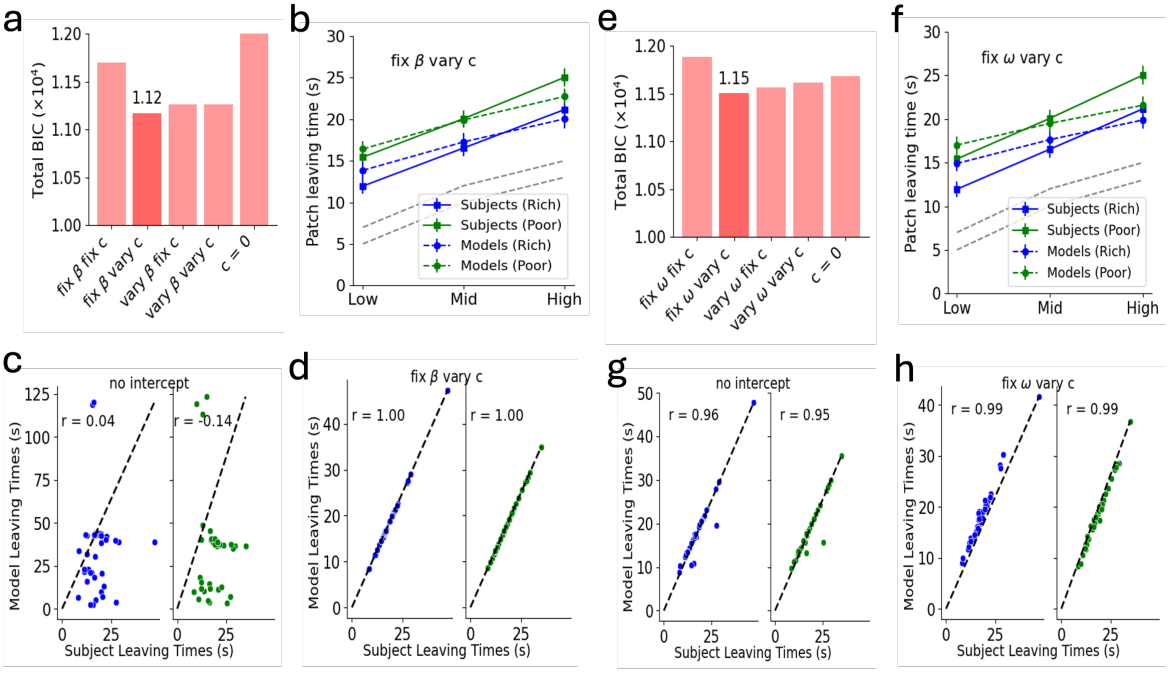

Figure 3.5: **Comparison of softmax and mellowmax parameter fits to the empirical data.** Panels (a)–(d) correspond to the softmax model, while (e)–(h) refer to the mellowmax model. Panels (a) and (e) show sum of BIC scores for various model-fitting scenarios. Panels (b) and (f) compare predicted and observed mean leaving times of the particpants in the Leheron's patch foraging experiment. Panels (c-h) plot predicted leaving times against empirical data, with intercepts enhancing models' ability to capture individual differences in (d) and (h).

Conversely, the mellowmax exhibited less sensitivity to the inclusion of an intercept (Figures 3.5g-h), suggesting that its adaptive ω parameter is sufficiently robust to capture individual variability. This adaptability, which allows the mellowmax to adjust its sensitivity to reward differences, likely explains its robustness in reflecting personal decision-making patterns [26]. However, this flexibility necessitates a complex optimization process to obtain the adaptive $\hat{\beta}$ at each decision-making step, rendering mellowmax computationally slower [57], albeit parameter-efficient.

3.3.2 Parameter Changes Driven by Environmental Context

To investigate how the softmax and mellowmax models adapt to various environments, we examined the fitted parameters β (softmax) and ω (mellowmax) in rich and poor environments. We focused on the case where the intercept was fixed, allowing β or ω to vary. One participant was excluded due to outlier behavior—a higher mean leaving time

in the rich environment compared to the poor, which contradicts predictions from MVT [6]. We considered p-values ≤ 0.05 as indicative of significant differences.

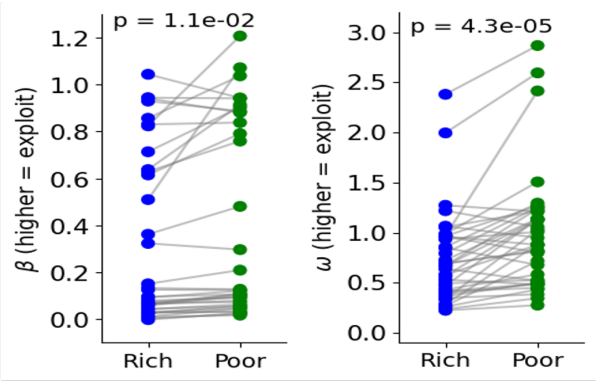

Figure 3.6: **Parameter fits across rich and poor environments for softmax (β) and mellowmax (ω) models.** The left plot shows β values for the softmax model, and the right shows ω values for the mellowmax model. Higher values in poorer environments indicate greater exploitation tendencies, supported by p-values.

As illustrated in Figure 3.6 (left panel), the β in the softmax model increased significantly in the poorer environment ($p = 0.011$), indicating a shift toward exploitation. This trend aligns with earlier findings (Figure 3.3) and optimal foraging theory, which suggests that individuals exploit resources more when they are scarce [5, 33].

Similarly, the mellowmax model showed a significant increase in ω in the poorer environment ($p = 4.3 \times 10^{-5}$), indicating a higher tendency toward exploitation when resources are limited. These findings are consistent with both theoretical predictions (Figure 3.4) and prior empirical studies [29, 26].

These results highlight how environmental context influences decision-making strategies, with that both β and ω increasing in poorer environments, while the *intercept* representing the forager's bias, remaining fixed across environments. This study also highlights the utility of parameter fitting in stochastic models for capturing the complexity of foraging behavior in dynamic contexts [9].

3.4 Analysis of Variability in Patch-Leaving Behaviour

Analyzing variability, particularly through the standard deviation of leaving times, provides valuable insights into the cognitive processes behind decision-making. It highlights both the consistency and flexibility of individual strategies in different contexts, which helps us understand how foragers adapt to unpredictable environments [29, 58]. This section examines the patterns of variability observed in a patch foraging task [15], assesses stochastic models' predictive power, and explores how uncertainty influences stochasticity in foraging decisions.

3.4.1 Variability Patterns Across Patch Types and Environments

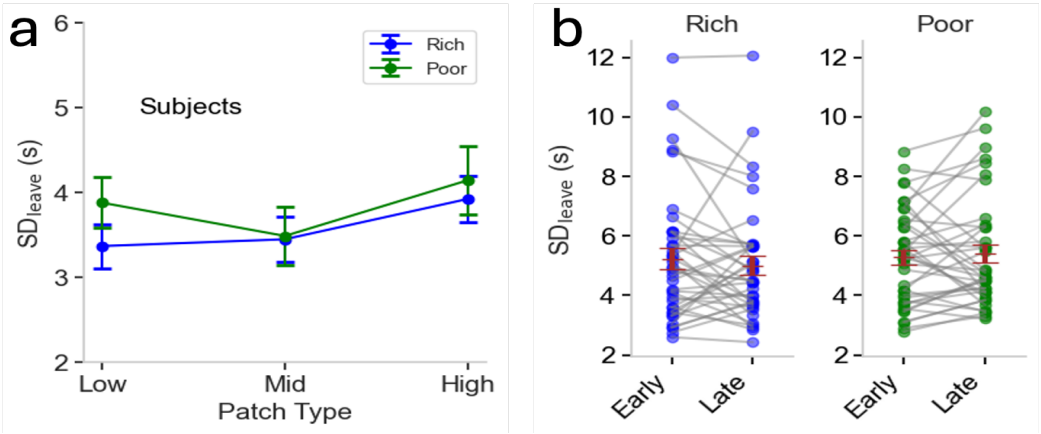

Figure 3.7: **Observed variability in leave times across subjects and environments.** (a) Line plots with error bars comparing the mean standard deviation of leave times across Low, Mid, and High patch types in rich and poor environments. (b) (b) Paired dot plots showing changes in standard deviation from early to late stages of the trial. Each grey line represents trends for a subject.

Figure 3.7a presents the variability in leaving times across patch types and environments from Le Heron's patch foraging experiment. Unlike mean leaving times, there is no clear pattern in the variability data, indicating that while average behaviors may be influenced by environmental factors, variability appears more random. This could reflect individual differences in decision-making, with some foragers using more exploratory approaches, contributing to increased variability [5, 58].

In Figure 3.7b, the lack of a significant reduction in variability over time challenges the expectation that experience would lead to more consistent behavior, reducing variance [30]. Instead, the persistent variability suggests that subjects did not converge on a stable strategy, likely due to ongoing uncertainty or the need to balance exploration and exploitation throughout the experiment [29, 3].

3.4.2 Predicting Stochasticity: Softmax vs. Mellowmax

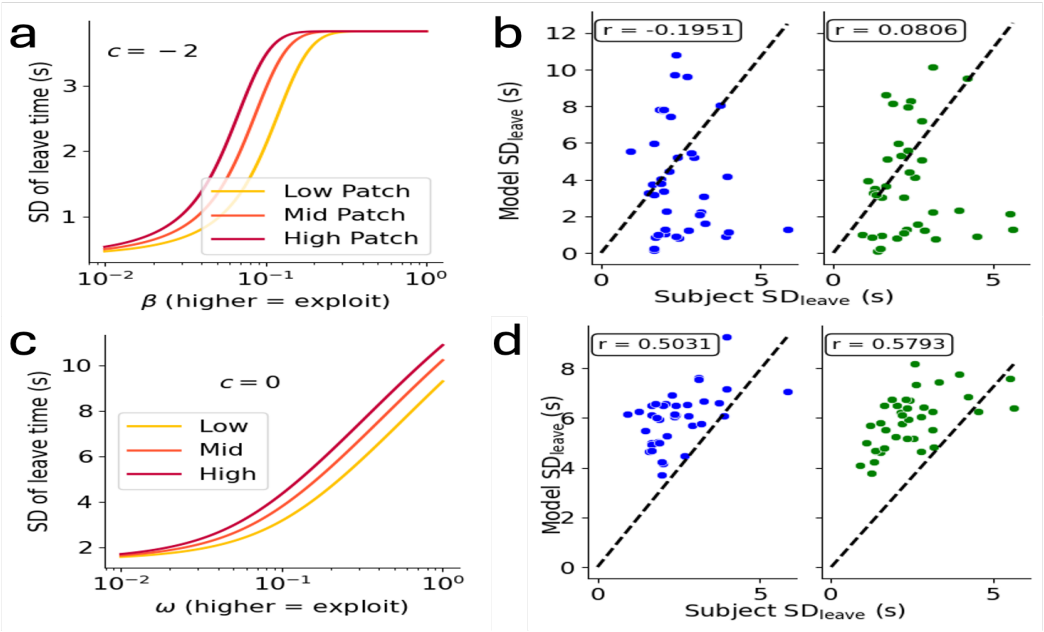

Figure 3.8: **Model predictions of variability in leave times.** (a) Standard deviation of leave times as a function of β (softmax) across patch types. (b) Scatter plots comparing predicted and observed standard deviations in leave times from the softmax model fit to each subject. (c) Standard deviation of leave times as a function of ω (mellowmax) across patch types. (d) (d) Scatter plots comparing predicted and observed standard deviations for the mellowmax model fit per subject. Mellowmax shows a higher correlation, suggesting a better fit for stochastic behavior.

We analyzed how well stochastic action selection models capture variability in leaving times by investigating the relationship between model parameters β or ω and their variability in leaving times. Figures 3.8a and 3.8c show the standard deviation of leaving times across different patch types as a function of model parameters. For the softmax model, variability increases with β before plateauing, indicating a stabilization of behavior as the model becomes more exploitative. The mellowmax model displays higher variability across the parameter range, highlighting its sensitivity.

Next, we compared model fits for each subject. Scatter plots in Figures 3.8b and 3.8d show predicted versus observed variability. The softmax model produced weaker correlations, implying it may struggle to fully capture the stochastic nature of behavior. Conversely, the mellowmax model, despite being slightly less accurate in predicting average leave times, showed higher correlation with observed variability, suggesting it better captures stochasticity.

These results indicate that while softmax is effective for modeling average behavior, mellowmax excels at accounting for both mean behavior and variability, making it more suited for dynamic decision-making tasks. This supports the notion that human foragers likely use flexible strategies balancing both mean behaviour and variability [20, 30].

The higher correlation for mellowmax fits (Figure 3.8d) also indicates a trade-off between optimizing mean and variability, aligning with studies on decision-making under uncertainty, which emphasize the importance of considering both average performance and variability in predicting human behavior [59, 60].

3.4.3 Understanding Stochasticity in Patch-Leaving Through Uncertainty

Understanding how agents manage uncertainty is crucial in analyzing decision-making processes, particularly in foraging tasks where decisions about when to leave a resource patch are vital. In such tasks, directed exploration helps agents navigate environments with uncertain reward rates, balancing between exploiting known resources and exploring new options to reduce uncertainty [3]. A key factor influencing these decisions is the decay parameter, which determines how rapidly rewards diminish in a patch. This study investigates whether uncertainty in this parameter impacts decision-making by introducing variability in the agents' behavior.

We simulated a foraging scenario based on Le Heron's experiment by assigning a normal distribution to the decay rate for each subject [15]. The variance of this distribution represented the agent's uncertainty about the decay rate. The agent sampled from this distribution to estimate the reward rate, introducing stochasticity into the decision process. As the agent sampled repeatedly, it refined its reward estimates for each patch, with the MVT guiding patch-leaving decisions (see Section 2.1.2).

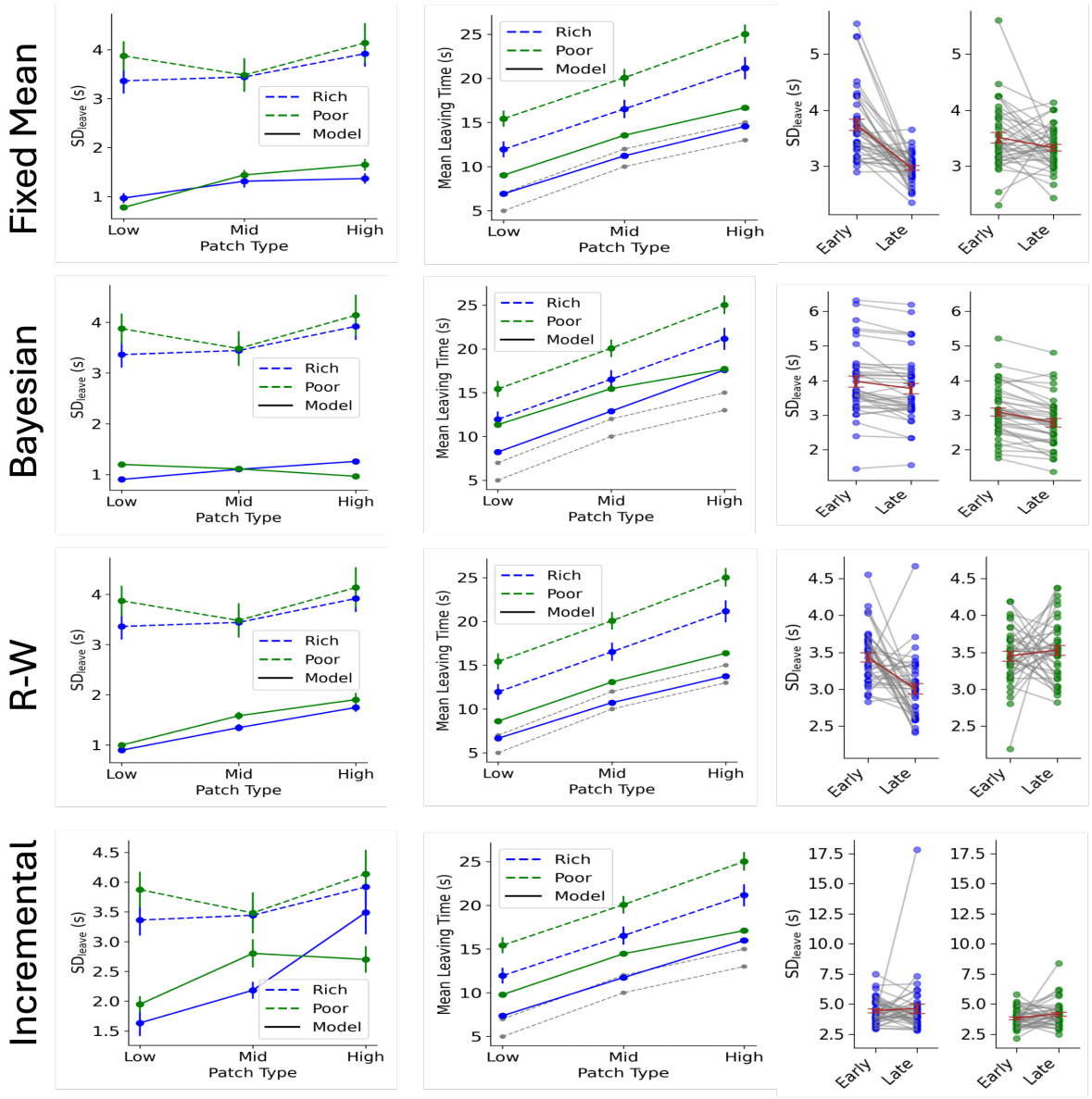

Figure 3.9: **Exploration and Uncertainty-Driven Stochasticity in Patch-Leaving Decisions.** The left column shows SD_{leave} for each method compared to empirical data. Variance-reducing approaches (top two rows) include a fixed mean strategy and Bayesian updating of both mean and variance, while fixed-variance methods (bottom two rows) use Rescorla-Wagner updates and incremental learning. The middle column compares expected leaving times between models and participants, while the right column shows changes in SD_{leave} from early to late experiment stages, with the mean highlighted in brown.

We applied belief-updating methods to separately adjust the mean and variance of the decay rate for each patch type and environment.To maintain stability, updates occurred

only when the agent left a patch. Initially, we used methods that reduced variance over trials (top two rows of Figure 3.9), including (1) a fixed mean strategy with decreasing variance and (2) Bayesian updating of both the mean and variance. These methods followed principles of directed exploration. We also examined fixed-variance approaches based on empirical findings, focusing on incremental updates to the mean (bottom two rows). These methods included (1) Rescorla-Wagner updates and (2) incremental learning based on the difference between the sampled and estimated decay rates

As shown in the left column of Figure 3.9, all methods predicted lower variability (SD_{leave}) than observed empirically, suggesting hyperparameter tuning is needed. The middle column shows simulated expected leaving times aligning with empirical data, particularly in capturing the patch and environmental effects. This alignment is attributed to the MVT-based decision rule, which effectively replicates general trends in human foraging behaviour [5, 29]. Additionally, lower variability may stem from shorter expected leaving times, as the mean-to-SD ratio remains consistent.

The right column shows how variability (SD_{leave}) changed from early to late stages of the experiment. Variance-reducing methods showed a general decline in variability, while fixed-variance methods exhibited no clear trends, indicating a lack of directed exploration. This suggests methods with variance updates may better predict directed exploration in foraging tasks. However, methods focusing solely on mean updates matched better with empirical data, supporting earlier findings that participants showed no learning trends (Figure 3.7). This contrasts with recent studies on uncertainty in decision-making, which suggest simpler heuristics may drive foraging decisions [3, 26]. These results suggest uncertainty in the decay parameter is not the primary driver of stochasticity in Leheron's experiment.

Chapter 4

Discussion

MVT is foundational to Optimal Foraging Theory, but human behavior frequently deviates from its predictions. While our study reaffirmed adaptive foraging in line with MVT, participants often chose to exploit known resources rather than explore new ones [20, 15]. These deviations may reflect adaptive strategies rather than inefficiencies. Recent research suggests that overharvesting may result from optimal learning and adaptation [61]. Neuroimaging studies also highlight the roles of the anterior cingulate cortex and hippocampus in these decisions [62, 63]. Furthermore, ancestral environments may shape foraging strategies, contributing to inconsistencies with MVT [64]. Incorporating predation risk into models [65], as suggested, enhances our understanding of foraging by accounting for the complexity of real-world environments.

Memory is critical to human foraging behavior. Our stochastic models, which focused on immediate rewards, overlooked the role of memory in real-world decision-making [66]. Research shows that spatial structure and memory significantly affect foraging strategies, with resource distribution influencing decision-making [67]. Human foraging is shaped by factors beyond immediate rewards, such as visual cues and abstract representations of large-scale environments [68]. Additionally, sex differences have been observed, with males typically persisting longer in patches than females [69]. Future models should incorporate these factors—memory, spatial context, and sex differences—to better capture the complexities of human foraging behavior.

Basic models cannot fully explain the complexity of human decision-making. Our findings show that individual variability in foraging behavior requires more than the basic softmax model. This supports the view that inter-individual differences must be considered in decision-making models [2]. Introducing a bias term into the softmax

operator improved predictions of leaving times across environments, highlighting the influence of ecological contexts [32]. Additionally, stochastic models tend to overfit expected values, neglecting behavioral variability [70]. Incorporating measures like standard deviation into model fitting may improve robustness in capturing the complexity of human decision-making [71].

Mellowmax excels in adaptive decision-making. The model performed well in capturing individual differences and predicting patch-leaving variability, aligning with insights from decision neuroscience, which highlights adaptive mechanisms in the brain [72, 73, 74]. Mellowmax's dynamic temperature adjustment may mirror neural processes involved in foraging decisions. Brain regions such as the frontal cortex may modulate decision parameters in response to changing environments [75, 76, 77]. Neural adaptations like these are also observed in honeybee foraging, where octopamine influences task engagement [78]. While mellowmax is not a direct map of neural circuits, its principles align with current models of neural optimization in complex environments [2]. Future research should investigate whether the brain uses mellowmax-like computations.

Participants did not demonstrate learning in the Le Heron patch foraging task. Behavioral variability, measured by standard deviation, remained stable across trials, contrasting with previous findings where learning tasks typically reduce variability as subjects adapt [3, 4]. Reduced variability is common in decision-making paradigms [8, 79]. The lack of learning in our study raises questions about whether task complexity, feedback, or other factors inhibited learning. Further research should explore whether this is specific to our experimental setup or a broader issue in foraging paradigms [29]. Future studies could investigate task modifications, such as adjusting reward structures or environmental cues, to better understand when learning occurs in patch foraging contexts [56].

The role of uncertainty in foraging decisions remains unclear. Our findings suggest that uncertainty in the decay parameter was inconclusive in explaining the stochasticity of patch-leaving times. This invites alternative approaches to understanding decision-making under uncertainty. Hierarchical Bayesian models, which account for uncertainty at both the patch and environment levels, may help explore exploration strategies in the task [80]. Such models could provide insights into whether human foragers favor directed exploration based on prior knowledge or random exploration characterized by stochastic behavior [25]. Bayesian updates could clarify the adaptive strategies employed by human foragers in uncertain environments.

Chapter 5

Conclusion

This study extends the understanding of human decision-making in foraging contexts by applying the experimental framework established by Le Heron et al. Our findings reaffirm the divergence of human behavior from Marginal Value Theorem (MVT) predictions, highlighting the limitations of deterministic models in real-world decisions. To address this, we assessed stochastic action-selection models—epsilon-greedy, softmax, and mellowmax. While the softmax model needed an added bias term to account for individual differences in exploration-exploitation, the mellowmax model, with adaptive parameters, provided a more consistent fit across subjects, even capturing variability in patch-leaving times. This supports mellowmax as a more generalized model for human foraging decisions in diverse settings. Uncertainty-based methods yielded inconclusive results regarding the role of decay parameters in patch-leaving decisions. Overall, our findings stress the importance of adaptive stochastic models like mellowmax that incorporate individual differences and environmental factors, contributing to decision neuroscience and enhancing theoretical foraging models with broader decision-making applications.

Bibliography

1. Daniel K. Thinking, fast and slow. 2017
2. Rangel A, Camerer C, and Montague PR. A framework for studying the neurobiology of value-based decision making. Nature reviews. Neuroscience 2008 Jun; 9:545–56. DOI: 10.1038/nrn2357. Available from: https://www.nature.com/articles/nrn2357
3. Cohen JD, McClure SM, and Yu AJ. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philosophical Transactions of the Royal Society B Biological Sciences 2007 Mar; 362:933–42. DOI: 10.1098/rstb.2007.2098. Available from: https://doi.org/10.1098/rstb.2007.2098
4. Daw ND, O'Doherty JP, Dayan P, Seymour B, and Dolan RJ. Cortical substrates for exploratory decisions in humans. Nature 2006 Jun; 441:876–9. DOI: 10.1038/nature04766. Available from: https://www.nature.com/articles/nature04766
5. Stephens DW and Krebs JR. Foraging theory. Vol. 6. Princeton university press, 1986
6. Charnov EL. Optimal foraging, the marginal value theorem. Theoretical Population Biology 1976 Apr; 9:129–36. DOI: 10.1016/0040-5809(76)90040-x. Available from: https://www.sciencedirect.com/science/article/abs/pii/004058097690040X
7. Hertwig R and Erev I. The description–experience gap in risky choice. Trends in Cognitive Sciences 2009 Dec; 13:517–23. DOI: 10.1016/j.tics.2009.09.004. Available from: https://doi.org/10.1016/j.tics.2009.09.004
8. Behrens TEJ, Woolrich MW, Walton ME, and Rushworth MFS. Learning the value of information in an uncertain world. Nature Neuroscience 2007 Aug; 10:1214–21. DOI: 10.1038/nn1954. Available from: https://doi.org/10.1038/nn1954

9. McNamara J. Optimal patch use in a stochastic environment. Theoretical Population Biology 1982 Apr; 21:269–88. DOI: 10.1016/0040-5809(82)90018-1. Available from: https://doi.org/10.1016/0040-5809(82)90018-1

10. Rushworth MF, Noonan MP, Boorman ED, Walton ME, and Behrens TE. Frontal cortex and Reward-Guided learning and Decision-Making. Neuron 2011 Jun; 70:1054–69. DOI: 10.1016/j.neuron.2011.05.014. Available from: https://doi.org/10.1016/j.neuron.2011.05.014

11. Hayden BY, Pearson JM, and Platt ML. Neuronal basis of sequential foraging decisions in a patchy environment. Nature Neuroscience 2011 Jun; 14:933–9. DOI: 10.1038/nn.2856. Available from: https://www.nature.com/articles/nn.2856

12. Kacelnik A, Vasconcelos M, Monteiro T, and Aw J. Darwin's "tug-of-war" vs. starlings' "horse-racing": how adaptations for sequential encounters drive simultaneous choice. Behavioral Ecology and Sociobiology 2010 Dec; 65:547–58. DOI: 10.1007/s00265-010-1101-2. Available from: https://doi.org/10.1007/s00265-010-1101-2

13. Krebs JR, Kacelnik A, and Taylor P. Test of optimal sampling by foraging great tits. Nature 1978 Sep; 275:27–31. DOI: 10.1038/275027a0. Available from: https://doi.org/10.1038/275027a0

14. Bernstein C, Kacelnik A, and Krebs JR. Individual decisions and the distribution of predators in a patchy environment. Journal of Animal Ecology 1988 Oct; 57:1007. DOI: 10.2307/5108. Available from: https://doi.org/10.2307/5108

15. Heron CL, Kolling N, Plant O, Kienast A, Janska R, Ang YS, Fallon S, Husain M, and Apps MA. Dopamine Modulates Dynamic Decision-Making during Foraging. Journal of Neuroscience 2020 May; 40:5273–82. DOI: 10.1523/jneurosci.2586-19.2020. Available from: https://doi.org/10.1523/jneurosci.2586-19.2020

16. Pearson JM, Watson KK, and Platt ML. Decision making: the Neuroethological turn. Neuron 2014 Jun; 82:950–65. DOI: 10.1016/j.neuron.2014.04.037. Available from: https://doi.org/10.1016/j.neuron.2014.04.037

17. Bidari S, Hady AE, Davidson JD, and Kilpatrick ZP. Stochastic dynamics of social patch foraging decisions. Physical Review Research 2022 Aug; 4. DOI: 10.1103/physrevresearch.4.033128. Available from: https://doi.org/10.1103/physrevresearch.4.033128

18. Wilson RC and Collins AG. Ten simple rules for the computational modeling of behavioral data. eLife 2019 Nov; 8. DOI: 10.7554/elife.49547. Available from: https://doi.org/10.7554/elife.49547

19. Schneider NA, Ballintyn B, Katz D, Lisman J, and Pi HJ. Parametric shift from rational to irrational decisions in mice. Scientific Reports 2021 Jan; 11. DOI: 10.1038/s41598-020-79949-w. Available from: https://www.nature.com/articles/s41598-020-79949-w

20. Mobbs D, Trimmer PC, Blumstein DT, and Dayan P. Foraging for foundations in decision neuroscience: insights from ethology. Nature reviews. Neuroscience 2018 May; 19:419–27. DOI: 10.1038/s41583-018-0010-7. Available from: https://doi.org/10.1038/s41583-018-0010-7

21. Gershman SJ and Daw ND. Reinforcement Learning and Episodic Memory in Humans and Animals: an Integrative framework. Annual Review of Psychology 2017 Jan; 68:101–28. DOI: 10.1146/annurev-psych-122414-033625. Available from: https://doi.org/10.1146/annurev-psych-122414-033625

22. Wittmann BC, Daw ND, Seymour B, and Dolan RJ. Striatal activity underlies Novelty-Based choice in humans. Neuron 2008 Jun; 58:967–73. DOI: 10.1016/j.neuron.2008.04.027. Available from: https://doi.org/10.1016/j.neuron.2008.04.027

23. Dubois M, Habicht J, Michely J, Moran R, Dolan RJ, and Hauser TU. Human complex exploration strategies are enriched by noradrenaline-modulated heuristics. eLife 2021 Jan; 10. DOI: 10.7554/elife.59907. Available from: https://doi.org/10.7554/elife.59907

24. Daw ND and Frank MJ. Reinforcement learning and higher level cognition: Introduction to special issue. Cognition 2009 Dec; 113:259–61. DOI: 10.1016/j.cognition.2009.09.005. Available from: https://doi.org/10.1016/j.cognition.2009.09.005

25. Wilson RC, Geana A, White JM, Ludvig EA, and Cohen JD. Humans use directed and random exploration to solve the explore–exploit dilemma. Journal of Experimental Psychology General 2014 Jan; 143:2074–81. DOI: 10.1037/a0038199. Available from: https://doi.org/10.1037/a0038199

26. Gershman SJ. Deconstructing the human algorithms for exploration. Cognition 2018 Apr; 173:34–42. DOI: 10.1016/j.cognition.2017.12.014. Available from: https://doi.org/10.1016/j.cognition.2017.12.014

27. Addicott MA, Pearson JM, Sweitzer MM, Barack DL, and Platt ML. A primer on foraging and the Explore/Exploit Trade-Off for Psychiatry research. Neuropsychopharmacology 2017 May; 42:1931–9. DOI: 10.1038/npp.2017.108. Available from: https://www.nature.com/articles/npp2017108

28. Rosati AG. Foraging Cognition: Reviving the Ecological Intelligence Hypothesis. Trends in Cognitive Sciences 2017 Sep; 21:691–702. DOI: 10.1016/j.tics.2017.05.011. Available from: https://doi.org/10.1016/j.tics.2017.05.011

29. Hills TT, Todd PM, Lazer D, Redish AD, and Couzin ID. Exploration versus exploitation in space, mind, and society. Trends in Cognitive Sciences 2015 Jan; 19:46–54. DOI: 10.1016/j.tics.2014.10.004. Available from: https://doi.org/10.1016/j.tics.2014.10.004

30. Constantino SM and Daw ND. Learning the opportunity cost of time in a patch-foraging task. Cognitive Affective Behavioral Neuroscience 2015 Apr; 15:837–53. DOI: 10.3758/s13415-015-0350-y. Available from: https://doi.org/10.3758/s13415-015-0350-y

31. Asadi K and Littman ML. An alternative softmax operator for reinforcement learning. *International Conference on Machine Learning*. PMLR. 2017 :243–52

32. Kolling N, Behrens TEJ, Mars RB, and Rushworth MFS. Neural mechanisms of foraging. Science 2012 Apr; 336:95–8. DOI: 10.1126/science.1216930. Available from: https://doi.org/10.1126/science.1216930

33. Pyke GH. Optimal foraging theory: a critical review. Annual review of ecology and systematics 1984; 15:523–75

34. Nonacs P and Soriano JL. Patch sampling behaviour and future foraging expectations in Argentine ants, Linepithema humile. Animal behaviour 1998; 55:519–27

35. Green RF. Optimal foraging for patchily distributed prey: random search. 1988

36. Thrun SB. Efficient exploration in reinforcement learning. Carnegie Mellon University, 1992

37. Sutton RS and Barto AG. Reinforcement learning: An introduction. MIT press, 2018

38. Kaelbling LP, Littman ML, and Moore AW. Reinforcement learning: A survey. Journal of artificial intelligence research 1996; 4:237–85

39. Auer P. Finite-time Analysis of the Multiarmed Bandit Problem. 2002

40. Press WH. Numerical recipes 3rd edition: The art of scientific computing. Cambridge university press, 2007

41. Lee D, Seo H, and Jung MW. Neural basis of reinforcement learning and decision making. Annual Review of Neuroscience 2012 Jul; 35:287–308. DOI: 10.1146/annurev-neuro-062111-150512. Available from: https://doi.org/10.1146/annurev-neuro-062111-150512

42. Nelder JA and Mead R. A simplex method for function minimization. The Computer Journal 1965 Jan; 7:308–13. DOI: 10.1093/comjnl/7.4.308. Available from: https://doi.org/10.1093/comjnl/7.4.308

43. Schwarz G. Estimating the dimension of a model. The annals of statistics 1978 :461–4

44. Burnham KP and Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research 2004; 33:261–304

45. Myung IJ. The importance of complexity in model selection. Journal of mathematical psychology 2000; 44:190–204

46. Tversky A and Kahneman D. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. science 1974; 185:1124–31

47. Tenenbaum JB, Griffiths TL, et al. The rational basis of representativeness. *Proceedings of the 23rd annual conference of the Cognitive Science Society*. Vol. 6. 2001

48. Rescorla RA. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. Classical conditioning, Current research and theory 1972; 2:64–9

49. Schulz E and Gershman SJ. The algorithmic architecture of exploration in the human brain. Current opinion in neurobiology 2019; 55:7–14

50. Hills TT, Jones MN, and Todd PM. Optimal foraging in semantic memory. Psychological review 2012; 119:431

51. Sims DW, Humphries NE, Hu N, Medan V, and Berni J. Optimal searching behaviour generated intrinsically by the central pattern generator for locomotion. eLife 2019 Nov; 8. DOI: 10.7554/elife.50316. Available from: https://doi.org/10.7554/elife.50316

52. Gabay AS and Apps MAJ. Foraging optimally in social neuroscience: computations and methodological considerations. Social Cognitive and Affective Neuroscience 2020 Mar; 16:782–94. DOI: 10.1093/scan/nsaa037. Available from: https://doi.org/10.1093/scan/nsaa037

53. Mehlhorn K, Newell BR, Todd PM, Lee MD, Morgan K, Braithwaite VA, Hausmann D, Fiedler K, and Gonzalez C. Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. Decision 2015; 2:191

54. Reverdy PB, Srivastava V, and Leonard NE. Modeling human decision making in generalized Gaussian multiarmed bandits. Proceedings of the IEEE 2014; 102:544–71

55. Daw ND et al. Trial-by-trial data analysis using computational models. Decision making, affect, and learning: Attention and performance XXIII 2011; 23:3–38

56. Hutchinson JMC, Wilke A, and Todd PM. Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches? Animal Behaviour 2008 Apr; 75:1331–49. DOI: 10.1016/j.anbehav.2007.09.006. Available from: https://doi.org/10.1016/j.anbehav.2007.09.006

57. Miahi E, MacQueen R, Ayoub A, Masoumzadeh A, and White M. Resmax: An Alternative Soft-Greedy Operator for Reinforcement Learning. Transactions on Machine Learning Research

58. McNamara JM, Green RF, and Olsson O. Bayes' theorem and its applications in animal behaviour. Oikos 2006 Jan; 112:243–51. DOI: 10.1111/j.0030-1299.2006.14228.x. Available from: https://doi.org/10.1111/j.0030-1299.2006.14228.x

59. Gershman SJ. Uncertainty and exploration. Decision 2019; 6:277

60. Wilson RC, Bonawitz E, Costa VD, and Ebitz RB. Balancing exploration and exploitation with information and randomization. Current opinion in behavioral sciences 2021; 38:49–56

61. Harhen NC and Bornstein AM. Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. Proceedings of the National Academy of Sciences 2023 Mar; 120. DOI: 10.1073/pnas.2216524120. Available from: https://doi.org/10.1073/pnas.2216524120

62. Shenhav A, Botvinick MM, and Cohen JD. The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 2013; 79:217–40

63. Kolling N, Wittmann MK, Behrens TE, Boorman ED, Mars RB, and Rushworth MF. Value, search, persistence and model updating in anterior cingulate cortex. Nature neuroscience 2016; 19:1280–5

64. Kaplan H. The evolution of the human life course. Between Zeus and the salmon: The biodemography of longevity 1997 :175–211

65. Lima SL. Back to the basics of anti-predatory vigilance: the group-size effect. Animal Behaviour 1995; 49:11–20

66. Vila J. Can we study episodic-like memory in preschoolers from an animal foraging model? APA PsycNet 2021; 10:123–30. Available from: https://psycnet.apa.org/buy/2021-92066-011

67. Kerster BE, Rhodes T, and Kello CT. Spatial memory in foraging games. Cognition 2016 Mar; 148:85–96. DOI: 10.1016/j.cognition.2015.12.015. Available from: https://doi.org/10.1016/j.cognition.2015.12.015

68. Crivelli-Decker J, Clarke A, Park SA, Huffman DJ, Boorman ED, and Ranganath C. Goal-oriented representations in the human hippocampus during planning and navigation. Nature Communications 2023; 14:141. DOI: 10.1038/s41467-023-35967-6. Available from: https://www.nature.com/articles/s41467-023-35967-6

69. Garcia M, Gupta S, and Wikenheiser AM. Sex differences in patch-leaving foraging decisions in rats. Oxford Open Neuroscience 2023; 2:1. DOI: 10.1093/oons/kvad011. Available from: https://doi.org/10.1093/oons/kvad011

70. Dayan P and Daw ND. Decision theory, reinforcement learning, and the brain. Cognitive Affective Behavioral Neuroscience 2008 Dec; 8:429–53. DOI: 10.3758/cabn.8.4.429. Available from: https://doi.org/10.3758/cabn.8.4.429

71. Pezzulo G, Rigoli F, and Friston KJ. Hierarchical active inference: a theory of motivated control. Trends in Cognitive Sciences 2018 Apr; 22:294–306. DOI: 10.1016/j.tics.2018.01.009. Available from: https://doi.org/10.1016/j.tics.2018.01.009

72. Platt M, Dayan P, Dehaene S, McCabe K, Menzel R, Phelps E, Plassmann H, Ratcliff R, Shadlen M, and Singer W. Neuronal correlates of decision making. 2008 May :125–54. DOI: 10.7551/mitpress/7735.003.0009. Available from: https://doi.org/10.7551/mitpress/7735.003.0009

73. Gold JI and Shadlen MN. The neural basis of decision making. Annual Review of Neuroscience 2007 Jul; 30:535–74. DOI: 10.1146/annurev.neuro.29.051605.113038. Available from: https://doi.org/10.1146/annurev.neuro.29.051605.113038

74. Cisek P. Cortical mechanisms of action selection: the affordance competition hypothesis. 2011 Nov :208–38. DOI: 10.1017/cbo9780511731525.015. Available from: https://www.cambridge.org/core/books/abs/modelling-natural-action-selection/cortical-mechanisms-of-action-selection-the-affordance-competition-hypothesis/CAE120F3CE5284D590C6EB505B289F54

75. Ding L and Gold JI. Neural Correlates of Perceptual Decision Making before, during, and after Decision Commitment in Monkey Frontal Eye Field. Cerebral Cortex 2011 Jul; 22:1052–67. DOI: 10.1093/cercor/bhr178. Available from: https://doi.org/10.1093/cercor/bhr178

76. Kable JW and Glimcher PW. The Neurobiology of Decision: Consensus and Controversy. Neuron 2009 Sep; 63:733–45. DOI: 10.1016/j.neuron.2009.09.003. Available from: https://doi.org/10.1016/j.neuron.2009.09.003

77. Heekeren HR, Marrett S, and Ungerleider LG. The neural systems that mediate human perceptual decision making. Nature reviews. Neuroscience 2008 May; 9:467–79. DOI: 10.1038/nrn2374. Available from: https://www.nature.com/articles/nrn2374

78. Farooqui T. Octopamine-Mediated Neuromodulation of insect senses. Neurochemical Research 2007 May; 32:1511–29. DOI: 10.1007/s11064-007-9344-7. Available from: https://doi.org/10.1007/s11064-007-9344-7

79. Badre D, Doll BBB, Long NM, and Frank MJ. Rostrolateral prefrontal cortex and individual differences in Uncertainty-Driven exploration. Neuron 2012 Feb; 73:595–607. DOI: 10.1016/j.neuron.2011.12.025. Available from: https://doi.org/10.1016/j.neuron.2011.12.025

80. McElreath R. Statistical rethinking. 2018 Jan. DOI: 10.1201/9781315372495. Available from: https://doi.org/10.1201/9781315372495