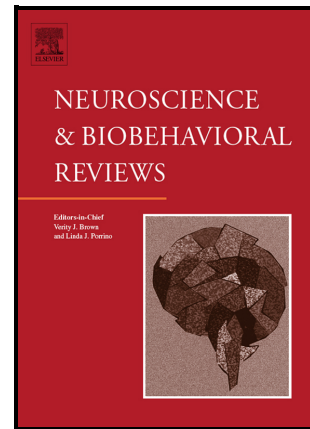


Journal Pre-proof

Hierarchical control over foraging behavior by anterior cingulate cortex

Ricardo J. Alejandro, Clay B. Holroyd



PII: S0149-7634(24)00092-7

DOI: <https://doi.org/10.1016/j.neubiorev.2024.105623>

Reference: NBR105623

To appear in: *Neuroscience and Biobehavioral Reviews*

Received date: 3 November 2023

Revised date: 14 February 2024

Accepted date: 13 March 2024

Please cite this article as: Ricardo J. Alejandro and Clay B. Holroyd, Hierarchical control over foraging behavior by anterior cingulate cortex, *Neuroscience and Biobehavioral Reviews*, (2024)
doi:<https://doi.org/10.1016/j.neubiorev.2024.105623>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier.

Hierarchical control over foraging behavior by anterior cingulate cortex

Ricardo J. Alejandro, Clay B. Holroyd

Department of Experimental Psychology, Ghent University, Ghent, Belgium

Corresponding author: Ricardo J. Alejandro, Department of Experimental Psychology, Ghent University, Ghent, Belgium; Ricardo.AlejandroBenavides@UGent.be; +32 (0)9 264 6407

Conflict of interest

The authors have indicated that they have no potential conflicts of interest to disclose.

Abstract

Foraging is a natural behavior that involves making sequential decisions to maximize rewards while minimizing the costs incurred when doing so. The prevalence of foraging across species suggests that a common brain computation underlies its implementation. Although anterior cingulate cortex is believed to contribute to foraging behavior, its specific role has been contentious, with predominant theories arguing either that it encodes environmental value or choice difficulty. Additionally, recent attempts to characterize foraging have taken place within the reinforcement learning framework, with increasingly complex models

scaling with task complexity. Here we review reinforcement learning foraging models, highlighting the hierarchical structure of many foraging problems. We extend this literature by proposing that ACC guides foraging according to principles of model-based hierarchical reinforcement learning. This idea holds that ACC function is organized hierarchically along a rostral-caudal gradient, with rostral structures monitoring the status and completion of high-level task goals (like finding food), and midcingulate structures overseeing the execution of task options (subgoals, like harvesting fruit) and lower-level actions (such as grabbing an apple).

Keywords: foraging, anterior cingulate cortex, ACC, hierarchical reinforcement learning, sequential decision making

Introduction

Real-world decisions often involve selecting a single choice from a clearly defined set of options, such as choosing an item from a menu. This type of choice has been the mainstay of laboratory studies of decision making for decades, but many everyday behaviors entail making sequential or continuous decisions, such as whether or not to continue watching the same TV show, or whether to accept or reject a series of job offers. Interest in this type of problem—when to accept, reject, and persist at extended behaviors—dates to the 1960's, when researchers in behavioral ecology and ethology applied insights from economics to explain how animals forage for resources in their environment (Stephens et al., 2007). Foraging constitutes an interesting research problem because it is common to most species and, in contrast to tasks that are typically conducted in the laboratory, involves naturalistic behaviors that sometimes require strong problem solving skills. The subject therefore appeals to researchers in disciplines beyond ecology, including evolution (Cisek, 2019), anthropology

(Smith et al., 1983), psychiatry (Addicott et al., 2017; Barack & Platt, 2017), social neuroscience (Contreras-Huerta et al., 2022; Gabay & Apps, 2021), marketing (Wells et al., 2023), logistics (O’Fallon et al., 2023), computer science (J. Wu & Aberer, 2003), artificial intelligence (Rathore et al., 2022), robotics (Winfield, 2009), behavioral psychology (Kamil, 1983; Pitkow & Angelaki, 2017) and cognitive neuroscience (Hayden & Walton, 2014).

Foraging occurs in natural environments, which tend to be highly complex. Resources are often sparsely clustered in sub-sections of the environment called “patches” (Charnov, 1976) that can vary in richness and effort demands. Animals often encounter patches that they never previously visited, which they can then exploit before moving on to other patches. Hence, the animals are confronted with a series of travel-harvest-decide (i.e., stay or leave) sequences of behavior (Stephens & Krebs, 1986). They must navigate these decisions with the twin goals of maximizing profit (such as food or money) and minimizing costs (including time, resources, effort, and energy) based on their history of rewards earned and costs incurred (Zhang et al., 2015).

Moreover, the problem is made even harder when the sources of reward deplete, when travelling between patches exacts a high cost (in terms of energy expenditure, time delays, etc.), and when the rewards are known only after being harvested. Although precise information about the foraging environment—such as the patches’ reward distributions, decay rates, the overall richness of the environment, and so on—could significantly facilitate these decisions, such information is rarely available to the forager in real-world scenarios. Hence, foragers typically infer this information by interacting with the environment itself. In that sense, they face a significant challenge: the foragers must actively *learn* properties of their environment in order to determine the optimal balance between when to exploit a patch versus when to explore for alternatives (the *exploration/exploitation dilemma*) (March, 1991; Radner & Rothschild, 1975; Sutton & Barto, 2018). These complexities make foraging a challenging

problem both for animals to solve and for researchers to understand. Note that, by design, foraging tasks studied in the laboratory are often much simpler and lack many of the characteristics of their counterparts in the wild. For example, animals are sometimes required to choose/switch between two available patches instead of making sequential choices between a current patch and the next, unknown patch. As we define the problem here, foraging studies involve sequences of choices between exploiting a current source of reward that depletes with use vs. switching to an alternative source of reward while incurring a cost because of the switch.

Although computational approaches related to evidence accumulation and Bayesian inference (Davidson & El Hady, 2019; Harhen & Bornstein, 2023; Kane et al., 2022; Kilpatrick et al., 2021) have elucidated aspects of it, foraging is fundamentally a reinforcement learning (RL) problem (Morimoto, 2019). In RL, a decision-making *agent* (be it a human, animal, or computational process) attempts to maximize long-term reward by evaluating the *value* of the current *state* of the environment and its available *actions*. The values are typically learned from the rewards and costs obtained by repeated interactions with the environment. Choices are then dictated by a *policy* that probabilistically maps the state transitions (Sutton & Barto, 2018). The computational power of this framework, together with natural parallels to the neural foundations of learning and behavior (Krausz et al., 2023; Schultz, 2013; Schultz et al., 1997), have put RL algorithms in the spotlight of the cognitive neuroscience of animal learning and decision-making. Crucially, RL principles can be used to model how agents learn new tasks from rewards and penalties in the absence of specific task instructions (Kaelbling et al., 1996), permitting inspection of the internal variables that drive behaviors such as foraging.

Further, insight into foraging can be gained by studying its underlying neural mechanisms. Notably, neuroimaging and neurophysiology studies have consistently implicated the anterior cingulate cortex (ACC) in such behavior (Hayden, Pearson, et al., 2011; Hayden & Walton, 2014; Kane et al., 2022; Kolling et al., 2012; Shenhav et al., 2014; Wittmann et al., 2016). The

ACC is a brain structure that extends along the frontal midline cortex and is believed to be involved in cognitive control and decision making, but its specific function is highly debated (Holroyd & Verguts, 2021). This controversy extends to ACC's role in foraging (Ebitz & Hayden, 2016), a behavior that ACC could support by way of RL signals communicated there from the midbrain dopamine system (Holroyd & Coles, 2002; see also Schultz et al., 1997; Schultz & Dickinson, 2000). This possibility is consistent with studies showing that pharmacological deactivation of dopamine signals impairs foraging behavior in rodents (F. Li, Cao, et al., 2012) and humans (Constantino et al., 2017; Le Heron et al., 2020; Rutledge et al., 2009), and that markers of dopamine synthesis and availability (as measured by positron emission tomography) correlate with patch leaving thresholds (Ianni et al., 2023).

Here we propose that complex foraging behaviors draw on specialized RL strategies including the abilities to plan prospectively and to adapt flexibly to environmental changes. We start by reviewing the strengths and weaknesses of current RL-based foraging theories, emphasizing how the need to add parameters to simpler models associated with so-called model-free RL results in more complex models that have more in common with so-called model-based RL. Then, we extend this research by proposing that many animals forage using a computational strategy called model-based hierarchical reinforcement learning (MB-HRL). In brief, we suggest that animals represent foraging tasks according to a hierarchical organization that renders the problem computationally tractable. By grouping contextually-related actions into goal-directed sequences, higher levels of the hierarchy are freed to solve complex foraging problems without being overwhelmed by details of task execution, which are delegated to the lower levels of the hierarchy. Further, model-based representations of the environment enable the forager to simulate and update goal-directed plans on-the-fly. We also highlight the ACC as an important neural structure orchestrating this mechanism, with hierarchically-higher representations implemented in more rostral parts of the ACC. In so doing, we argue that the

key contributions of ACC to foraging illuminate the computational underpinnings of foraging behavior itself.

Marginal Value Theorem

Foraging concerns the choices that animals make when serially encountering patches of resources within an environment. Each encounter confronts them with a decision: either to harvest from the current patch or to switch to an alternative, potentially better patch in the environment (Charnov & Orians, 1973; Stephens & Krebs, 1986). Given a finite amount of resources per patch, repeated harvests from the same patch usually results in depletion, which compels the forager eventually to search elsewhere. In addition, both the act of harvesting rewards as well as travelling to new patches can incur costs in terms of time and energy. Therefore, to maximize their returns, foragers must optimize resource expenditure by staying in each patch only when doing so remains profitable, and by moving between patches only when necessary. The main computational challenge thus becomes to find the optimal moment to abandon a patch. Foraging studies have shown that birds, rodents, non-human- and human primates, and other species (Freidin & Kacelnik, 2011; Gabay & Apps, 2021; Hayden, Pearson, et al., 2011; Kacelnik, 1984; Kane et al., 2017; Wolfe, 2013), as well as plants (McNickle & Cahill, 2009) and artificial neural network agents (Wispiński et al., 2023), all make patch-leaving decisions according to the Marginal Value Theorem (MVT). The MVT holds that the optimal moment to abandon a patch is when the reward rate of the current patch drops below the current average reward rate across all other patches in the environment (Charnov, 1976; Stephens & Krebs, 1986) (Figure 1).

The success of the MVT over the past decades has established it as the flagship theory of foraging research. This suggests that an average reward rate solution for solving decision-making problems in general (Dayan, 2009; Niv et al., 2007) might also apply to foraging

problems in particular. In fact, the MVT has been proven to approach optimality in cases where the reward rate of the patches decreases monotonically with repeated harvesting (Kolling & Akam, 2017; McNamara & Houston, 1985). In such scenarios, the MVT's average reward threshold rule is seen to explain patch-leaving decisions better than cumulative long-term reward rules like temporal difference (TD) learning (see below) (Constantino & Daw, 2015).

That said, several strong assumptions underlying the MVT framework limit its wider applicability to more general foraging problems (Kilpatrick et al., 2020). For example, the MVT assumes that reward rates decay monotonically with repeated patch engagement, and MVT approaches optimality in scenarios that meet this constraint, but this is unrealistic in some natural settings where reward availability is highly volatile. In particular, competing foragers can cause the resource to deplete suddenly (Slotow & Coumi, 2000), and patches can yield stochastic returns and/or improve over time. The MVT also assumes complete knowledge of resource availability on the part of the agent, which is rarely possible in the real world (Pirolli, 2007). Predators can also block access to the resources (Lima, 1998), and changes in environmental conditions can redistribute or eliminate the rewards (Peat & Goulson, 2005). The average reward rate also does not provide an accurate estimate of reward value for environments with highly uneven and variable sources of reward, as some patches can be richer than other patches, and some regions within a patch can be richer than other regions. As a result, the MVT fails to explain many real-world foraging problems (Kolling & Akam, 2017; Mobbs et al., 2018). Nevertheless, the parsimony of the MVT algorithm together with its ability to account for a wide variety of observations has established it as the benchmark in foraging research.

Model-free RL models of foraging.

An increasingly popular approach in recent years has been to describe foraging behavior according to principles of RL, which overcome some of the limitations of MVT models (Kolling & Akam, 2017). The first RL algorithms, which are collectively termed model-free (MF) RL, were derived from psychology studies of learning by reinforcement (Ferster & Skinner, 1957; Skinner, 1965; Thorndike, 1898). An agent following MF-RL maximizes reward by learning what actions to take via a series of trial-and-error interactions with the environment. In particular, temporal difference (TD) learning entails comparing whether the reward value of ongoing events is better or worse than the agent's predictions of those events, generating a *reward prediction error* (RPE) when predictions and outcomes disagree. In turn, these RPEs are used to update the predictive values of those events, with the objective of minimizing the absolute value of the RPE. This is achieved by associating earlier and earlier states or actions in a temporal sequence with more accurate predictive values (Sutton & Barto, 2018). Consequently, an MF-RL agent adaptively adjusts its behavior based on its improved predictions.

This approach highlights an important distinction between standard laboratory *bandit task* experiments, which have been used previously to examine MF-RL, and typical foraging problems (Averbeck, 2015). In bandit tasks, an agent must repeatedly select a bandit among a set of alternatives, after which a reward is delivered according to a probability specific to that bandit. By contrast, foragers often encounter only one patch at a time, and therefore the value of the alternative patches remain unknown until the next patch is selected. Further, the switch from one patch to another typically incurs a cost (Hayden, 2018). Viewed as a bandit task problem, this would translate as a repeated choice between the bandit at hand and the averaged return of all other bandits, with a penalty paid when the current bandit is abandoned to choose one of the others. Hence, applications of RL principles to foraging problems must incorporate an inherent cost associated with exploring alternatives, in addition to any harvesting costs. In

particular, an MF-RL foraging agent should estimate both the value of the current offer and the average value of the environment based on RPE-driven changes in reward value (Pereira-Obilinovic et al., 2022), in order to estimate short-term reward values of upcoming harvests to guide patch-leaving decisions (Figure 1).

Constantino and Daw (2015) investigated whether two types of MF-RL algorithms, TD(λ) and R-learning, would account for patch-leaving behavior better than the MVT approach. They utilized a task that required human participants to harvest fruit from trees in different virtual environments. Each harvest incurred a short time delay (a *harvest cost*), and moving to a new tree incurred a longer time delay (a *travel cost*); systematic variation of the travel costs and reward depletion rates across blocks produced environments with differing degrees of richness. Importantly, the reward rate for each tree depleted monotonically throughout the experiment (i.e., decreased consistently without significant reversals or fluctuations). They observed that an MVT-based model better predicted participant patch-leaving decisions than did the other models, highlighting the superiority of the average reward rate rule used in MVT relative to a long-term discounted cumulative rewards algorithm used in MF-RL, at least for environments with patches characterized by monotonically depleting rewards.

By contrast, the MVT is less accurate for non-deterministic (and more realistic) environments that are characterized by non-monotonically changing reward rates. For example, in a task where the richness of the environments changed over time, participants learned from positive outcomes during periods of patch improvement faster than they did from negative outcomes during periods of patch deterioration; this adaptive behavior was better explained by an RL model that incorporated separate learning rates for positive and negative outcomes, which does not have an equivalent in the standard MVT implementation (Garrett & Daw, 2020). Further, MF-RL algorithms that incorporate such a bias—i.e., that differentially weight the different outcomes (with beneficial outcomes having greater influence over behavior than non-beneficial

outcomes) (Garrett & Daw, 2020)—account better than MVT models for a tendency of foragers to remain in a patch for longer than optimal, which has been observed across multiple empirical studies (Constantino & Daw, 2015; Harhen et al., 2021; Harhen & Bornstein, 2022; Hutchinson et al., 2008; Lenow et al., 2017; Wikenheiser et al., 2013). More generally, this *overharvesting* phenomenon has been related to a range of cognitive, computational, and personality-related processes such as temporal perception adjustment (Brunner et al., 1992; Kendall & Wikenheiser, 2022), environment uncertainty (Kilpatrick et al., 2021), environment structure learning (Harhen & Bornstein, 2023), risk sensitivity (Constantino & Daw, 2015), error compensation (Cash-Padgett & Hayden, 2020), and apathy and compulsivity traits (Scholl et al., 2022). Considered in this wider context, the prevalence of overharvesting is a salient feature of foraging that merits explanation.

That said, standard MF-RL models struggle to capture many behaviors of real foragers unless the models incorporate additional parameters or features. In one study that tested humans in a virtual reality foraging environment, an MF-RL model did not accurately simulate their observed foraging trajectories between patches in the environment. Only by incorporating prior knowledge into the model in the form of initial policy values (which were estimated via imitation learning, whereby the artificial agent is trained by imitating example behaviors; Pomerleau, 1991) did the model account well for the empirical observations (Giammarino et al., 2022). In practice, not only are foraging environments often non-stationary and volatile, but actions at any point in time can influence the availability of subsequent choices (Calhoun & Hayden, 2015), which render the learning problem more difficult for MF-RL to solve (Gershman & Daw, 2017; Sutton & Barto, 2018).

Moreover, MF-RL models fail to account for several key features of natural foraging behavior, such as memory of resource locations (Applegate & Aronov, 2022) and flexible long-term planning (Drummond & Niv, 2020) that animals exhibit in natural environments (Fagan et al.,

2013). Relatedly, MF-RL models cannot predict patch replenishment while the patches are not being harvested, which often directs patch (re)visiting of foragers in the real world (Hall-McMaster & Luyckx, 2019). Animals can also balance the benefits of exploitation and exploration by remembering the locations of resources according to their predictability (Jackson et al., 2020). In fact, a recent study suggested that mice might arbitrate between model-free and model-based behavioral strategies in response to their degree of experience with the foraging task (Le et al., 2022). These task demands require the agent to adapt their decisions according to the latest outcomes and to integrate detailed information about the environment (Shteingart & Loewenstein, 2014). Importantly, this information is not incorporated in models that compare the average reward rate across patches to the latest received reward (MVT models; Figure 1), nor in models that predict trial-by-trial outcomes driven by latest updates in patch reward value (MF-RL models; Figure 1). Thus, despite similarities between foraging and the (often simple) RL rules that are commonly studied in cognitive psychology, it seems likely that the neurocognitive mechanisms underlying foraging benefit from incorporating more complex representations of the environment.

Model-based RL models of foraging

In general, adding parameters to relatively simple RL algorithms improves their ability to capture important features of behavior but also increases their complexity. In so doing, these models tend to progressively resemble model-based RL algorithms that leverage detailed statistical structure about environmental relationships. In contrast to MF-RL, model-based (MB) RL provides a means for agents to predict the reward value of sequences of decisions and direct behavior accordingly. The agent first learns an internal model of the environment that maps the probabilities that each state will transition to every other state, and then applies this model to simulate the consequences of its decisions (Botvinick & Weinstein, 2014). In this way, MB-RL can enable planning over extended sequences of behaviors (Figure 2), which in

the case of foraging addresses some of the shortcomings of the RL solutions discussed above. In fact, structural knowledge (i.e., a model) of the environment might be a requirement for navigating the complex dynamics of naturalistic tasks (Noel et al., 2021). Given that during foraging each choice influences subsequent states, and such consequences could be distributed over time, a complete description of environmental state transition probabilities would enable optimal foraging behavior even in non-stationary environments (Davidson & El Hady, 2019; Harhen et al., 2021; Harhen & Bornstein, 2022; Kanarek et al., 2008; McNamara, 1982). Consistent with these task demands, a growing body of research suggests that human behavior relies, at least in part, on MB-RL (Daw & Dayan, 2014; Gershman & Niv, 2015; Keramati et al., 2016; Pouncy et al., 2021; but see Collins & Cockburn, 2020; Momennejad, 2020; Reid & Staddon, 1998), though its role in foraging is only recently being examined (Kumar et al., 2019; Noel et al., 2021; Yoo et al., 2020).

MB-RL thus provides a potentially powerful means to address difficult foraging problems. However, MB-RL is limited by the computational intractability of learning a complete internal model capable of simulating all possible world states, at least for the large and complex environments that are typically encountered by foragers in the real world. Instead, agents can sometimes learn an approximate internal model of the environment (Song et al., 2016), though the complexity of these models still scales with the complexity of the foraging problem (Doll et al., 2012). For example, in a study with human participants where the reward value of different patches either improved or worsened over time, an RL model that detected trends in short- and long-term reward rates explained foraging choices better than a standard MF-RL model did (Wittmann et al., 2016). The model learned long-term estimates of RPEs by introducing an *expected RPE* parameter into an RL algorithm. Working in parallel with a reward value learning RL algorithm, this feature enabled the model to forecast changing reward rates to guide decision-making. The model accounted well for empirically-observed decisions

to stay in a patch, which were positively influenced by recent outcomes and negatively influenced by outcomes farther in the past. Hence, the increased model complexity ensured success even in scenarios with high variability.

Beyond the fact that real-world patches sometimes return better or worse harvests over time, they also sometimes replenish when not visited: apples grow back on trees, shelves are restocked in supermarkets, and so on. Animals have been observed to remember and leverage this kind of information when selecting patches (Merkle et al., 2014). As well, when provided with opportunities to re-visit patches from one trial to the next, human participants evaluate both global information related to average reward rate of the environment and local information related to patch-specific reward and repletion rates. In such cases, foragers obtain more reward by exploiting patches that replenish faster, which they learn about from prior experience. This behavior is well-represented by a model that includes estimates of both current reward value and reward replenishment rate for each alternative patch, in addition to the average reward value estimate of the environment (Hall-McMaster et al., 2021).

As mentioned, large state/action spaces can present a computational challenge to model-based RL algorithms that simulate planned behaviors. Many RL implementations of foraging can be stymied by environments where the number of possible states and actions increases non-linearly with the number of steps in a behavioral sequence (Zhang et al., 2015), and where environmental features such as patch richness can be heterogeneous and volatile. One way in which the dimensionality of the state space can be reduced is via artificial neural networks that are trained to generalize across similar states (Colin et al., 2023). For instance, Wu et al. (2020) reduce the dimensionality by assuming that the agent's internal model is encoded within a low-dimensional state manifold. They trained a recurrent neural network agent to solve a foraging task (Sugrue et al., 2004), which they used to estimate the likelihood of possible models of the

environment dynamics (i.e., the agent's internal model) given the agent's choices (Z. Wu et al., 2020).

In summary, although MB-RL provides flexibility that enables the agent to respond to unexpected changes in reward contingencies, the planning process is computationally expensive for large state spaces (Gershman, 2017). In practice, this results in long training times and slow simulation times that, in humans, demand neurocognitive resources associated with high cognitive effort (Otto, Gershman, et al., 2013; Otto, Raio, et al., 2013).

Foraging as a hierarchical task

The computational demands of model-based planning in complex environments can be further ameliorated by breaking down the problem into smaller, hierarchically-organized components. Hierarchical reinforcement learning (HRL) can minimize the computational load associated with many MB-RL problems by grouping representations of states and actions according to their implicit hierarchical relationships (Barto & Mahadevan, 2003; Hutsebaut-Buysse et al., 2022). In particular, a common implementation of HRL relies on the concept of *options* (Sutton et al., 1999), which are goal-directed action policies with specific initiation and termination states. Given that many behaviors naturally exhibit a hierarchical organization (Cooper & Shallice, 2000; Jeon, 2014; Lashley, 1951; G. A. Miller et al., 1960), the options framework can allow for decomposing temporally extended, complex behaviors into nested chains of action policies corresponding to tasks, options, and actions (Botvinick et al., 2009; Sutton et al., 1999). Doing so makes difficult learning problems more tractable by reducing the dimensionality of the problem space. In particular, grouping together low-level actions according to common subgoals, and in turn grouping these according to the main goal, reduces the size of the state spaces for learning and planning for each level of the hierarchy, thereby increasing the system's computational efficiency (Botvinick et al., 2009).

Similar to its *flat* counterparts, HRL can be implemented using model-free (MF-HRL) or model-based (MB-HRL) approaches, or both. In these cases, the MF- or MB-RL computations are executed across every level of the hierarchy (Figure 2), allowing for decisions at different levels to be driven either by the cached value of options or by the inferred value of extended plans that simulate the execution of sequences of options (termed *saltatory HRL*) (Botvinick & Weinstein, 2014). Thus, MB-HRL can reduce the computational expense of planning in complex environments by simulating behavioral sequences at a high level of temporal abstraction, rather than by simulating the low level sequences of actions that comprise the policies. Importantly, MB-HRL can facilitate planning even in scenarios where the agent does not have access to a complete model of the environment (Pateria et al., 2022), as progress towards the goal can be simulated and monitored by the higher-levels even if the lower level policies are incomplete (Levy et al., 2017; Singh et al., 2004).

The temporal and state abstractions associated with HRL mainly address value-based decision problems that are characterized by large state and action spaces (Eckstein & Collins, 2020). Specifically, as a complex value-based learning problem, foraging tasks can be simplified using abstract hierarchical representations of the planning space that map actions to options, thereby reducing the computational toll of model-based representations (Chalmers et al., 2016). A reasonable way to describe foraging problems according to principles of MB-HRL is by assuming that the agent's primary goal is to maximize food intake, with each foraging patch (such as a forest) represented as a *sub-option*. In turn, each sub-option can be described as a set of contextually-related individual actions (like approaching a tree, grabbing a tree branch, pulling fruit, and so on) operating under a single option policy according to a common subgoal (such as picking fruit) (Figure 1). In contrast, with flat representations, all the actions and decisions operate under the same task policy and are given the same significance, such that the forager must learn to select actions and make decisions based solely on the current state (not

considering broader goals or subgoals). By utilizing hierarchical representations, rather than planning across each individual action, foraging agents can plan and predict harvests and patch-switching decisions based on how profitable the current patch is relative to previously experienced, alternative patches. And based on the profitability of the current option (driven by the aggregate profit being earned from the option's sub-options) relative to the overall profit from the entire foraging task, the highest level of the model can then direct switches between options (e.g., from fishing to picking fruit, and vice versa) as appropriate.

Options can also optimize computational demands when existing subroutines can be adapted to novel situations with similar high-level contexts (Hengst, 2011). For instance, a forager trying to maximize the amount of food intake could collect apples using the *fruit-picking* policy. When encountering a pear tree, the fruit-picking policy can be reused and adapted to collect pears as well, as both activities fall within the domain of the high-level policy.

In summary, because many foraging problems can be naturally decomposed into tasks with hierarchical structure, incorporating hierarchically-organized representations into existing foraging architectures can optimize computational resources by reducing the decision and planning space. Organized thusly, the higher levels can plan over and monitor progress toward the main goal at a relatively simple level of abstraction, and instigate strategy changes over subgoals according to the task demands. Hierarchy further provides modular compartmentalization that enables generalization across tasks and subtasks without interference between submodules (Colin et al., 2023; Holroyd & Verguts, 2021).

Foraging and the anterior cingulate cortex

The computational basis of foraging can be elucidated by understanding its underlying brain mechanisms. From this perspective, the ACC looms especially large. ACC activity has been observed ubiquitously in a wide variety of neuroimaging studies, which indicates that it

contributes to behavior more broadly than only foraging (Heilbronner & Hayden, 2016). For example, the ACC has been associated with conflict monitoring (Botvinick et al., 2001), error likelihood estimation (Brown & Braver, 2005), action selection (Holroyd & Coles, 2002), effort motivation (Holroyd & McClure, 2015), value estimation (Scholl et al., 2015), reward volatility estimation (Behrens et al., 2007), effort avoidance (Botvinick, 2007), attention and surprise (Hayden, Heilbronner, et al., 2011), effort allocation (Verguts et al., 2015), decision costs evaluation (Crosson et al., 2009), and many other processes (Rushworth et al., 2011). As a consequence of this complexity, the function of ACC is highly disputed (Holroyd & Verguts, 2021), but our limited understanding of ACC can still provide insight into its contribution to foraging.

Multiple brain areas track changes in environmental- and decision-related variables during foraging (Bari et al., 2019; Cazettes et al., 2023; Rudebeck & Izquierdo, 2021; Shahidi et al., 2019; Silston et al., 2021; Sugrue et al., 2004), but ACC is consistently observed to be engaged in foraging experiments (Pearson et al., 2014). Further, deactivating ACC seems to reliably impair foraging behavior (Kane et al., 2022; F. Li, Li, et al., 2012; Seamans et al., 1995; Tervo et al., 2021; Verтеchi et al., 2020), indicating that ACC plays an essential role in foraging. In an early investigation of foraging neuroscience, Hayden, Pearson, et al. (2011) found that the firing rates of individual neurons recorded from the monkey ACC increased with each consecutive decision to stay in a patch, until the neural activity crossed a threshold that predicted when that patch would be abandoned in favor of an alternative. This patch-leaving threshold was modulated by the travel cost (the travel time from the current patch to the next), suggesting that ACC neurons might be sensitive to the value of the environment, namely the tradeoff between the costs and benefits for both the chosen and unchosen choices. Further, a study that used an accept/reject paradigm (i.e., accept the current offer or reject it to continue the search) found that firing rates of neurons in monkey ACC encoded the reward magnitude

of the non-selected choice, consistent with a role for the ACC in keeping track of environmental value for the adaptive adjustment of behavioral strategies (Blanchard & Hayden, 2014).

A recent study examined the foraging behavior of mice in a task requiring switches between two different sites to obtain rewards (Vertechi et al., 2020). Optogenetic inactivation of ACC caused the mice to stay longer in a patch, evidently by altering the perceived value of the current patch against that of the environment (i.e., the alternative patch). Similarly, Tervo et al., (2021) proposed that switching to the alternative involves two different processes mediated by separate neural systems: a decision to leave the current patch and a subsequent decision to commit to the alternative. To test that hypothesis, they optogenetically manipulated ACC activity at two different times in rats performing a foraging task: on each trial when the animals were first presented with a choice between accepting or rejecting the current offer, and then following unrewarded feedback after a decision to accept (i.e., stay with the current offer). Whereas ACC perturbation during choice delayed switching to the alternative patch, ACC perturbation following unrewarded feedback increased the number of switches to the alternative (i.e., the opposite effect), which suggests that ACC might contribute differentially to foraging decisions. To test this speculation, they then perturbed two ACC output pathways—ACC intra-telencephalic and pyramidal tract pathways—and showed that these opposing effects on switching were independently mediated by these two neural groups.

Relatedly, a different neurophysiology study in rats found that individual ACC neuron activity correlated with foraging decision variables such as local (patch) and global (environment) reward rates, the value of leaving a patch, and decision difficulty (i.e., the similarity in the values of staying versus leaving a patch), whereas ACC activity averaged across neurons tracked the value of leaving a patch (Kane et al., 2022). When ACC was pharmacologically perturbed via a GABA receptor agonist, the animals tended to stay in the patches for more trials and their response times increased. The investigators attributed this effect to movement

slowness due to ACC inactivation, concluding that even though the ACC indeed encodes decision variables related to foraging, it plays a more general role in performance monitoring for the purpose of regulating response vigor (see also Niv et al., 2007).

Neuroimaging studies with human participants have also been revealing. One fMRI study required human participants to track the reward value of patches in a dynamic environment where the patches had increasing or decreasing reward rates. Computational simulations based on RL principles revealed that the ACC BOLD response tracked trends in reward rates that in turn were associated with the observed foraging decisions (Wittmann et al., 2016). Moreover, ACC activity was sensitive to trends in both long- and short-term past rewards, which exhibited negative and positive influences, respectively, on decisions to remain in a patch. In another fMRI study, Kolling et al. (2012) observed that ACC activation correlated with the value of exploring the alternative choice, suggesting that the ACC encodes the value of the environment. Similar findings were reported in a study using sequential choices from multiple alternatives, where ACC activations were said to be encoding adaptive decisions away from the default choice (Boorman et al., 2013). In another human fMRI study, Shenhav et al. (2014) reported increased ACC activation when two alternatives have similar estimated values, making the choice between them more difficult. They argued that the human ACC encodes the difficulty of the choice. And in a follow-up study, Kolling et al. (2018) showed that ACC activity increased with increased value of the environment, and found that although choice difficulty also drove ACC activity, this occurred only late in the trial. The authors argued that the value of the environment activates ACC early in the trial, but when decisions are difficult, the choice takes longer and consequently ACC activity remains high for longer. The diverging interpretations between these two studies sparked a subsequent debate about whether ACC contributes to foraging by comparing the value of the current choice against the value of alternative choices available in the environment, or by evaluating the conflict between

especially difficult choices (Ebitz & Hayden, 2016; Kolling, Wittmann, et al., 2016; Shenhav et al., 2016).

Foraging, ACC, and hierarchy

We suggest that the incorporation of a hierarchical structure into RL models of foraging can improve our understanding of its underlying neuro-computational mechanisms. In line with a widely-discussed proposal that hierarchical representations are organized according to a rostro-caudal gradient of abstraction along prefrontal cortex (Badre, 2008; Jeon, 2014; Koechlin et al., 2003; Koechlin & Summerfield, 2007; O'Reilly, 2010), including frontal midline cortex (Alexander & Brown, 2015; Taren et al., 2011), the representations underlying foraging may also be spatially distributed. In particular, one proposal based on principles of HRL holds that ACC adaptively selects and maintains task options (Holroyd & McClure, 2015; Holroyd & Verguts, 2021; Holroyd & Yeung, 2012). On this account, the most rostral areas of ACC (rACC) select and maintain high-level goals and options, anterior midcingulate cortex (aMCC, also called caudal or dorsal ACC) supports the execution of lower-level sub-options that implement the subgoals, and posterior midcingulate cortex (pMCC) and nearby cortical regions select the individual actions driven by the option-specific policies. Options and sub-options are selected via cortico-striatal feedback loops through the basal ganglia in accordance with changes in the average reward value at each hierarchical level. When the received reward is less than expected for that level, control increases over the immediately lower level in order to enhance task performance. When the received reward is equal or more than expected, control decreases over the immediately lower level, permitting that level to disengage gradually from task execution, thereby maximizing reward without mobilizing unnecessary effort.

A hierarchical structure in combination with value-driven learning aligns naturally with the organization of foraging problems. This view holds that rACC implements and sustains the

highest-level goal in foraging (e.g., finding food). When the overall task value is less than predicted, control increases over the immediately lower level (i.e., the option level), motivating a switch of task strategy (e.g., from fishing to hunting) which is reflected in increased activation of rACC. Lower in the hierarchy, aMCC implements sub-options related to the individual task subgoals (such as selecting individual patches for fishing or hunting). Reduced option value (as occurs when a patch is exhausted) would cause aMCC to increase control over the immediately lower level (implemented by pMCC), which is reflected in increased activation of aMCC and persistent behavior (Figure 3). Crucially, this account predicts that the activations in aMCC and rACC are anti-correlated, with increased activation of the former associated with persistence on the present task and increased activation of the latter associated with switches between tasks.

In addition, ACC appears to implement hierarchical control with model-based planning. In an MB-RL setting, the values associated with choices are assigned based on prospective simulations of the states and outcomes that such choices lead to. Converging with this idea, foraging-related studies highlight ACC as a candidate structure supporting model-based behavior. For example, a study with a two-step task in mice (option selection followed by action selection) found that ACC neurons encoded future states and reward transition probabilities, consistent with MB-RL, and that this trait was impaired after optogenetic manipulation of ACC (Akam et al., 2021). Further, a study in rats showed preferential and abrupt shifts in activations of neural populations within ACC depending on the animal's exploratory or exploitative state (Caracheo et al., 2013), in line with the flexibility afforded by model-based planning. In a competitive foraging setting, neuron activity in rat ACC represented potential competitive effort, particularly when such effort leads to high rewards (Hillman & Bilkey, 2012); and individual neuron activity in mice ACC encoded resource availability and social information such as rank within the group, while the ensemble neuron

activity encoded history of rewards and prospective competitive success (S. W. Li et al., 2022). Both these accounts suggest a role for ACC related to a competition/benefit analysis when evaluating prospective courses of action. And still another study in monkeys showed that, whereas neuronal spiking activity in ACC reflected rewards and cost information, local field potentials from the same area reflected only reward signals (Ramakrishnan et al., 2019), suggesting that ACC might integrate various sources of contextual information related to a cost-benefit arbitration into a single prospective decision variable.

These observations suggest that ACC represents states or task contexts (Rushworth et al., 2012) in order to facilitate planning (K. J. Miller & Venditto, 2021). In general, the complex response profiles observed in ACC are consistent with distributed coding of predictive world models (Kolling, Behrens, et al., 2016) across ACC neural ensembles (Shahnazian & Holroyd, 2018) that, in turn, appear to sub-serve hierarchically-organized, goal-directed action sequences (Holroyd et al., 2018). In particular, it has been argued that the control function of ACC may be especially important for persisting at the execution of high-level plans when the lower-level models are incomplete or unsuited for the task at hand (Holroyd & Verguts, 2021).

This perspective speaks also to the debate on the role of ACC in foraging (Ebitz & Hayden, 2016). Rodent studies have shown that disrupting the ACC tends to alter foraging decisions, and most importantly that the nature of these alterations depends on the location of the ACC manipulation. Consistent with its putative role in task switching, deactivation of rACC mainly leads to delayed switches to the alternative patch (Kane et al., 2022; Vertechchi et al., 2020). And consistent with the proposal that aMCC supports the execution of lower-level sub-options that implement task subgoals, inactivation of a border area between cingulate and midcingulate cortices affects model updating while preserving direct action-reward associations (Akam et al., 2021), which suggests roles in model-based planning and option evaluation. Further, in line with the MB-HRL framework, which proposes distributed responsibilities in ACC for

regulating exploiting vs. switching between patches, ACC continuously estimates and tracks the values of the patch and of the environment (Hayden, Pearson, et al., 2011; Kolling et al., 2012) in a spatially distributed and seemingly opposing manner (Tervo et al., 2021). Finally, the MB-HRL framework predicts the existence of hierarchically-organized control signals, such that conflict resulting from choices of comparable value is resolved by control signals applied by the immediately higher level, consistent with the choice difficulty explanation for foraging-related ACC activity (Shenhav et al., 2014). Needless to say, these cross-species generalizations should be interpreted with caution given controversy about homologs of human ACC in non-human animals (Cole et al., 2009; van Heukelum et al., 2020).

In this review we have emphasized RL descriptions of foraging. It is important to remark that the distinction between model-based RL vs model-free RL is not actually a dichotomy and in fact masks numerous subtleties related to animal behavior (Collins & Cockburn, 2020). Rather than being characterized by one or the other, decision problems take a variety of forms that align better with model-free representations, model-based representations, a combination of both, or entirely different algorithms such as latent-state inference (Harhen et al., 2021; Kumar et al., 2017), Bayesian inference (Harhen & Bornstein, 2023), and successor representations (Dayan, 1993; Gershman, 2018). However, we believe that the natural parallels of HRL with the hierarchical structure of foraging behavior and its neural implementation can provide exceptional insight into the phenomenon.

Conclusions and prospective work

Interest in foraging research has grown significantly in recent years. Foraging studies enable researchers to move beyond relatively controlled approaches for studying decision making (such as forced binary choices and go/no go paradigms) by providing ecologically valid experimental designs with complexity commensurate to the real-world problems that our brains

evolved to solve (Pretelli et al., 2022; Wise et al., 2023). At the same time, the use of a variety of computational models to explain foraging behavior is providing valuable new tools for theory development and testing. That said, exactly how foragers process information to make patch-leaving decisions, and how those decisions are computed in the brain, remains uncertain. Here we have highlighted important contributions of MF- and MB-RL foraging models, and have argued that the incorporation of hierarchically organized representations that regulate top-down control can improve on these models even further. Moreover, these computations appear to be implemented by ACC, which integrates the advantages of model-based planning and hierarchically structured behavior to facilitate foraging according to principles of MB-HRL. Collectively, the neural studies and modelling approaches reviewed here provide insight into the computational processes underlying foraging behavior.

Acknowledgements

This work is part of a project that has received funding from the European Research Council (ERC) under the EU's Horizon 2020 Research and Innovation Programme (grant agreement no. 787307).

References

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A

Primer on Foraging and the Explore/Exploit Trade-Off for Psychiatry Research.

Neuropsychopharmacology, 42(10), 1931–1939.

<https://doi.org/10.1038/npp.2017.108>

Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R. F., Dayan, P.,

& Costa, R. M. (2021). The Anterior Cingulate Cortex Predicts Future States to

Mediate Model-Based Action Selection. *Neuron*, 109(1), 149-163.e7.

<https://doi.org/10.1016/j.neuron.2020.10.013>

- Alexander, W. H., & Brown, J. W. (2015). Hierarchical Error Representation: A Computational Model of Anterior Cingulate and Dorsolateral Prefrontal Cortex. *Neural Computation*, 27(11), 2354–2410. https://doi.org/10.1162/NECO_a_00779
- Applegate, M. C., & Aronov, D. (2022). Flexible use of memory by food-caching birds. *eLife*, 11, e70600. <https://doi.org/10.7554/eLife.70600>
- Averbeck, B. B. (2015). Theory of Choice in Bandit, Information Sampling and Foraging Tasks. *PLOS Computational Biology*, 11(3), e1004164. <https://doi.org/10.1371/journal.pcbi.1004164>
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, 12(5), 193–200. <https://doi.org/10.1016/j.tics.2008.02.004>
- Barack, D. L., & Platt, M. L. (2017). Engaging and Exploring: Cortical Circuits for Adaptive Foraging Decisions. In J. R. Stevens (Ed.), *Impulsivity* (Vol. 64, pp. 163–199). Springer International Publishing. https://doi.org/10.1007/978-3-319-51721-6_6
- Bari, B. A., Grossman, C. D., Lubin, E. E., Rajagopalan, A. E., Cressy, J. I., & Cohen, J. Y. (2019). Stable Representations of Decision Variables for Flexible Behavior. *Neuron*, 103(5), 922–933.e7. <https://doi.org/10.1016/j.neuron.2019.06.001>
- Barto, A. G., & Mahadevan, S. (2003). Recent Advances in Hierarchical Reinforcement Learning. *Discrete Event Dynamic Systems*, 13(4), 341–379. <https://doi.org/10.1023/A:1025696116075>
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9), 1214–1221. <https://doi.org/10.1038/nn1954>

- Blanchard, T. C., & Hayden, B. Y. (2014). Neurons in Dorsal Anterior Cingulate Cortex Signal Postdecisional Variables in a Foraging Task. *Journal of Neuroscience*, *34*(2), 646–655. <https://doi.org/10.1523/JNEUROSCI.3151-13.2014>
- Boorman, E. D., Rushworth, M. F., & Behrens, T. E. (2013). Ventromedial Prefrontal and Anterior Cingulate Cortex Adopt Choice and Default Reference Frames during Sequential Multi-Alternative Choice. *The Journal of Neuroscience*, *33*(6), 2242–2253. <https://doi.org/10.1523/JNEUROSCI.3022-12.2013>
- Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(4), 356–366. <https://doi.org/10.3758/CABN.7.4.356>
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. <https://doi.org/10.1037/0033-295X.108.3.624>
- Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*(3), 262–280. <https://doi.org/10.1016/j.cognition.2008.08.011>
- Botvinick, M. M., & Weinstein, A. (2014). Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130480. <https://doi.org/10.1098/rstb.2013.0480>
- Brown, J. W., & Braver, T. S. (2005). Learned Predictions of Error Likelihood in the Anterior Cingulate Cortex. *Science*, *307*(5712), 1118–1121. <https://doi.org/10.1126/science.1105783>
- Brunner, D., Kacelnik, A., & Gibbon, J. (1992). Optimal foraging and timing processes in the starling, *Sturnus vulgaris*: Effect of inter-capture interval. *Animal Behaviour*, *44*(4), 597–613. [https://doi.org/10.1016/S0003-3472\(05\)80289-1](https://doi.org/10.1016/S0003-3472(05)80289-1)

- Calhoun, A. J., & Hayden, B. Y. (2015). The foraging brain. *Current Opinion in Behavioral Sciences*, 5, 24–31. <https://doi.org/10.1016/j.cobeha.2015.07.003>
- Caracheo, B. F., Emberly, E., Hadizadeh, S., Hyman, J. M., & Seamans, J. K. (2013). Abrupt changes in the patterns and complexity of anterior cingulate cortex activity when food is introduced into an environment. *Frontiers in Neuroscience*, 7. <https://doi.org/10.3389/fnins.2013.00074>
- Cash-Padgett, T., & Hayden, B. (2020). Behavioural variability contributes to over-staying in patchy foraging. *Biology Letters*, 16(3), 20190915. <https://doi.org/10.1098/rsbl.2019.0915>
- Cazettes, F., Mazzucato, L., Murakami, M., Morais, J. P., Augusto, E., Renart, A., & Mainen, Z. F. (2023). A reservoir of foraging decision variables in the mouse brain. *Nature Neuroscience*, 26(5), 840–849. <https://doi.org/10.1038/s41593-023-01305-8>
- Chalmers, E., Luczak, A., & Gruber, A. J. (2016). Computational Properties of the Hippocampus Increase the Efficiency of Goal-Directed Foraging through Hierarchical Reinforcement Learning. *Frontiers in Computational Neuroscience*, 10. <https://doi.org/10.3389/fncom.2016.00128>
- Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, 9(2), 129–136. [https://doi.org/10.1016/0040-5809\(76\)90040-X](https://doi.org/10.1016/0040-5809(76)90040-X)
- Charnov, E. L., & Orians, G. H. (1973). *Optimal Foraging: Some Theoretical Explorations*.
- Cisek, P. (2019). Resynthesizing behavior through phylogenetic refinement. *Attention, Perception, & Psychophysics*, 81(7), 2265–2287. <https://doi.org/10.3758/s13414-019-01760-1>
- Cole, M. W., Yeung, N., Freiwald, W. A., & Botvinick, M. (2009). Cingulate cortex: Diverging data from humans and monkeys. *Trends in Neurosciences*, 32(11), 566–574. <https://doi.org/10.1016/j.tins.2009.07.001>

- Colin, T. R., Ikkink, I., & Holroyd, C. B. (2023). *Distributed representations for cognitive control in frontal medial cortex* [Preprint]. <https://doi.org/10.1101/2023.12.12.571242>
- Collins, A. G. E., & Cockburn, J. (2020). Beyond dichotomies in reinforcement learning. *Nature Reviews Neuroscience*, *21*(10), 576–586. <https://doi.org/10.1038/s41583-020-0355-6>
- Constantino, S. M., Dalrymple, J., Gilbert, R. W., Varanese, S., Di Rocco, A., & Daw, N. D. (2017). *A Neural Mechanism for the Opportunity Cost of Time* [Preprint]. *Neuroscience*. <https://doi.org/10.1101/173443>
- Constantino, S. M., & Daw, N. D. (2015). Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(4), 837–853. <https://doi.org/10.3758/s13415-015-0350-y>
- Contreras-Huerta, L. S., Pisauro, M. A., Kuechenhoff, S., Gekiere, A., Le Heron, C., Lockwood, P., & Apps, M. A. J. (2022). *A reward self-bias leads to more optimal foraging for ourselves than others* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/8r45z>
- Cooper, R., & Shallice, T. (2000). CONTENTION SCHEDULING AND THE CONTROL OF ROUTINE ACTIVITIES. *Cognitive Neuropsychology*, *17*(4), 297–338. <https://doi.org/10.1080/026432900380427>
- Crosson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E. J., & Rushworth, M. F. S. (2009). Effort-Based Cost–Benefit Valuation and the Human Brain. *The Journal of Neuroscience*, *29*(14), 4531–4541. <https://doi.org/10.1523/JNEUROSCI.4515-08.2009>
- Davidson, J. D., & El Hady, A. (2019). Foraging as an evidence accumulation process. *PLOS Computational Biology*, *15*(7), e1007060. <https://doi.org/10.1371/journal.pcbi.1007060>

- Daw, N. D., & Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), 20130478. <https://doi.org/10.1098/rstb.2013.0478>
- Dayan, P. (1993). Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Computation*, 5(4), 613–624. <https://doi.org/10.1162/neco.1993.5.4.613>
- Dayan, P. (2009). Prospective and retrospective temporal difference learning. *Network: Computation in Neural Systems*, 20(1), 32–46. <https://doi.org/10.1080/09548980902759086>
- Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, 22(6), 1075–1081. <https://doi.org/10.1016/j.conb.2012.08.003>
- Drummond, N., & Niv, Y. (2020). Model-based decision making and model-free learning. *Current Biology*, 30(15), R860–R865. <https://doi.org/10.1016/j.cub.2020.06.051>
- Ebitz, R. B., & Hayden, B. Y. (2016). Dorsal anterior cingulate: A Rorschach test for cognitive neuroscience. *Nature Neuroscience*, 19(10), 1278–1279. <https://doi.org/10.1038/nn.4387>
- Eckstein, M. K., & Collins, A. G. E. (2020). Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences*, 117(47), 29381–29389. <https://doi.org/10.1073/pnas.1912330117>
- Fagan, W. F., Lewis, M. A., Auger-Méthé, M., Avgar, T., Benhamou, S., Breed, G., LaDage, L., Schlägel, U. E., Tang, W., Papastamatiou, Y. P., Forester, J., & Mueller, T. (2013). Spatial memory and animal movement. *Ecology Letters*, 16(10), 1316–1329. <https://doi.org/10.1111/ele.12165>

- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Appleton-Century-Crofts. <https://doi.org/10.1037/10627-000>
- Freidin, E., & Kacelnik, A. (2011). Rational Choice, Context Dependence, and the Value of Information in European Starlings (*Sturnus vulgaris*). *Science*, *334*(6058), 1000–1002. <https://doi.org/10.1126/science.1209626>
- Gabay, A. S., & Apps, M. A. J. (2021). Foraging optimally in social neuroscience: Computations and methodological considerations. *Social Cognitive and Affective Neuroscience*, *16*(8), 782–794. <https://doi.org/10.1093/scan/nsaa037>
- Garrett, N., & Daw, N. D. (2020). Biased belief updating and suboptimal choice in foraging decisions. *Nature Communications*, *11*(1), 3417. <https://doi.org/10.1038/s41467-020-16964-5>
- Gershman, S. J. (2017). Reinforcement Learning and Causal Models. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (pp. 295–306). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199399550.013.20>
- Gershman, S. J. (2018). The Successor Representation: Its Computational Logic and Neural Substrates. *The Journal of Neuroscience*, *38*(33), 7193–7200. <https://doi.org/10.1523/JNEUROSCI.0151-18.2018>
- Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*(1), 101–128. <https://doi.org/10.1146/annurev-psych-122414-033625>
- Gershman, S. J., & Niv, Y. (2015). Novelty and Inductive Generalization in Human Reinforcement Learning. *Topics in Cognitive Science*, *7*(3), 391–415. <https://doi.org/10.1111/tops.12138>
- Giammarino, V., Dunne, M. F., Moore, K. N., Hasselmo, M. E., Stern, C. E., & Paschalidis, I. Ch. (2022). *Combining imitation and deep reinforcement learning to accomplish*

human-level performance on a virtual foraging task.

<https://doi.org/10.48550/ARXIV.2203.06250>

Hall-McMaster, S., Dayan, P., & Schuck, N. W. (2021). Control over patch encounters changes foraging behavior. *iScience*, *24*(9), 103005.

<https://doi.org/10.1016/j.isci.2021.103005>

Hall-McMaster, S., & Luyckx, F. (2019). Revisiting foraging approaches in neuroscience.

Cognitive, Affective, & Behavioral Neuroscience, *19*(2), 225–230.

<https://doi.org/10.3758/s13415-018-00682-z>

Harhen, N. C., & Bornstein, A. M. (2022). Humans adapt their foraging strategies and computations to environment complexity. *Proceedings of the 5th Multidisciplinary Conference on Reinforcement Learning and Decision Making*, *5*.

Harhen, N. C., & Bornstein, A. M. (2023). Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. *Proceedings of the National Academy of Sciences*, *120*(13), e2216524120.

<https://doi.org/10.1073/pnas.2216524120>

Harhen, N. C., Hartley, C. A., & Bornstein, A. (2021). Model-based foraging using latent-cause inference. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. <https://escholarship.org/uc/item/9c33b182>

Hayden, B. Y. (2018). Economic choice: The foraging perspective. *Current Opinion in Behavioral Sciences*, *24*, 1–6. <https://doi.org/10.1016/j.cobeha.2017.12.002>

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., & Platt, M. L. (2011). Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. *The Journal of Neuroscience*, *31*(11), 4178–4187.

<https://doi.org/10.1523/JNEUROSCI.4652-10.2011>

- Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, *14*(7), 933–939.
<https://doi.org/10.1038/nn.2856>
- Hayden, B. Y., & Walton, M. E. (2014). Neuroscience of foraging. *Frontiers in Neuroscience*, *8*. <https://doi.org/10.3389/fnins.2014.00081>
- Heilbronner, S. R., & Hayden, B. Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annual Review of Neuroscience*, *39*(1), 149–170.
<https://doi.org/10.1146/annurev-neuro-070815-013952>
- Hengst, B. (2011). Hierarchical Reinforcement Learning. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 495–502). Springer US.
https://doi.org/10.1007/978-0-387-30164-8_363
- Hillman, K. L., & Bilkey, D. K. (2012). Neural encoding of competitive effort in the anterior cingulate cortex. *Nature Neuroscience*, *15*(9), 1290–1297.
<https://doi.org/10.1038/nn.3187>
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. <https://doi.org/10.1037/0033-295X.109.4.679>
- Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, *122*(1), 54–83. <https://doi.org/10.1037/a0038339>
- Holroyd, C. B., Ribas-Fernandes, J. J. F., Shahnazian, D., Silvetti, M., & Verguts, T. (2018). Human midcingulate cortex encodes distributed representations of task progress. *Proceedings of the National Academy of Sciences*, *115*(25), 6398–6403.
<https://doi.org/10.1073/pnas.1803650115>

- Holroyd, C. B., & Verguts, T. (2021). The Best Laid Plans: Computational Principles of Anterior Cingulate Cortex. *Trends in Cognitive Sciences*, 25(4), 316–329.
<https://doi.org/10.1016/j.tics.2021.01.008>
- Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, 16(2), 122–128.
<https://doi.org/10.1016/j.tics.2011.12.008>
- Hutchinson, J. M. C., Wilke, A., & Todd, P. M. (2008). Patch leaving in humans: Can a generalist adapt its rules to dispersal of items across patches? *Animal Behaviour*, 75(4), 1331–1349. <https://doi.org/10.1016/j.anbehav.2007.09.006>
- Hutsebaut-Buysse, M., Mets, K., & Latré, S. (2022). Hierarchical Reinforcement Learning: A Survey and Open Research Challenges. *Machine Learning and Knowledge Extraction*, 4(1), 172–221. <https://doi.org/10.3390/make4010009>
- Ianni, A. M., Eisenberg, D. P., Boorman, E. D., Constantino, S. M., Hegarty, C. E., Gregory, M. D., Masdeu, J. C., Kohn, P. D., Behrens, T. E., & Berman, K. F. (2023). PET-measured human dopamine synthesis capacity and receptor availability predict trading rewards and time-costs during foraging. *Nature Communications*, 14(1), 6122.
<https://doi.org/10.1038/s41467-023-41897-0>
- Jackson, B. J., Fatima, G. L., Oh, S., & Gire, D. H. (2020). Many Paths to the Same Goal: Balancing Exploration and Exploitation during Probabilistic Route Planning. *Eneuro*, 7(3), ENEURO.0536-19.2020. <https://doi.org/10.1523/ENEURO.0536-19.2020>
- Jeon, H.-A. (2014). Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. *Frontiers in Systems Neuroscience*, 8.
<https://doi.org/10.3389/fnsys.2014.00223>

- Kacelnik, A. (1984). Central Place Foraging in Starlings (*Sturnus vulgaris*). I. Patch Residence Time. *The Journal of Animal Ecology*, *53*(1), 283.
<https://doi.org/10.2307/4357>
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.
<https://doi.org/10.1613/jair.301>
- Kamil, A. C. (1983). Optimal Foraging Theory and the Psychology of Learning. *American Zoologist*, *23*(2), 291–302. <https://doi.org/10.1093/icb/23.2.291>
- Kanarek, A. R., Lamberson, R. H., & Black, J. M. (2008). AN INDIVIDUAL-BASED MODEL FOR TRADITIONAL FORAGING BEHAVIOR: INVESTIGATING EFFECTS OF ENVIRONMENTAL FLUCTUATION: AN INDIVIDUAL-BASED FORAGING MODEL. *Natural Resource Modeling*, *21*(1), 93–116.
<https://doi.org/10.1111/j.1939-7445.2008.00002.x>
- Kane, G. A., James, M. H., Shenhav, A., Daw, N. D., Cohen, J. D., & Aston-Jones, G. (2022). Rat Anterior Cingulate Cortex Continuously Signals Decision Variables in a Patch Foraging Task. *The Journal of Neuroscience*, *42*(29), 5730–5744.
<https://doi.org/10.1523/JNEUROSCI.1940-21.2022>
- Kane, G. A., Vazey, E. M., Wilson, R. C., Shenhav, A., Daw, N. D., Aston-Jones, G., & Cohen, J. D. (2017). Increased locus coeruleus tonic activity causes disengagement from a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(6), 1073–1083. <https://doi.org/10.3758/s13415-017-0531-y>
- Kendall, R. K., & Wikenheiser, A. M. (2022). Quitting while you're ahead: Patch foraging and temporal cognition. *Behavioral Neuroscience*, *136*(5), 467–478.
<https://doi.org/10.1037/bne0000526>

- Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proceedings of the National Academy of Sciences*, *113*(45), 12868–12873.
<https://doi.org/10.1073/pnas.1609094113>
- Kilpatrick, Z. P., Davidson, J. D., & El Hady, A. (2021). Uncertainty drives deviations in normative foraging decision strategies. *Journal of The Royal Society Interface*, *18*(180), 20210337. <https://doi.org/10.1098/rsif.2021.0337>
- Kilpatrick, Z. P., Davidson, J. D., & Hady, A. E. (2020). *Normative theory of patch foraging decisions* [Preprint]. *Animal Behavior and Cognition*.
<https://doi.org/10.1101/2020.04.22.055558>
- Koechlin, E., Ody, C., & Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, *302*(5648), 1181–1185.
<https://doi.org/10.1126/science.1088545>
- Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*(6), 229–235.
<https://doi.org/10.1016/j.tics.2007.04.005>
- Kolling, N., & Akam, T. (2017). (Reinforcement?) Learning to forage optimally. *Current Opinion in Neurobiology*, *46*, 162–169. <https://doi.org/10.1016/j.conb.2017.08.008>
- Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science*, *336*(6077), 95–98.
<https://doi.org/10.1126/science.1216930>
- Kolling, N., Behrens, T., Wittmann, M., & Rushworth, M. (2016). Multiple signals in anterior cingulate cortex. *Current Opinion in Neurobiology*, *37*, 36–43.
<https://doi.org/10.1016/j.conb.2015.12.007>

- Kolling, N., Scholl, J., Chekroud, A., Trier, H. A., & Rushworth, M. F. S. (2018). Prospection, Perseverance, and Insight in Sequential Behavior. *Neuron*, *99*(5), 1069-1082.e7. <https://doi.org/10.1016/j.neuron.2018.08.018>
- Kolling, N., Wittmann, M. K., Behrens, T. E. J., Boorman, E. D., Mars, R. B., & Rushworth, M. F. S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience*, *19*(10), 1280–1285. <https://doi.org/10.1038/nn.4382>
- Krausz, T. A., Comrie, A. E., Kahn, A. E., Frank, L. M., Daw, N. D., & Berke, J. D. (2023). Dual credit assignment processes underlie dopamine signals in a complex spatial environment. *Neuron*, *111*(21), 3465-3478.e7. <https://doi.org/10.1016/j.neuron.2023.07.017>
- Kumar, A., Wu, Z., Pitkow, X., & Schrater, P. (2019). Belief dynamics extraction. *CogSci - Annual Conference of the Cognitive Science Society.*, 2019, 2058–2064.
- Kumar, A., Wu, Z., Shahibi, N., Dragoi, V., Pitkow, X., & Schrater, P. (2017). Inferring latent states from foraging behavior. *Conference on Cognitive Computational Neuroscience*. Conference on Cognitive Computational Neuroscience.
- Lashley, K. S. (1951). The Problem of serial order in behavior. In *Cerebral mechanisms in behavior: The Hixon symposium* (pp. 112–136). Wiley.
- Le Heron, C., Kolling, N., Plant, O., Kienast, A., Janska, R., Ang, Y.-S., Fallon, S., Husain, M., & Apps, M. A. J. (2020). Dopamine Modulates Dynamic Decision-Making during Foraging. *The Journal of Neuroscience*, *40*(27), 5273–5282. <https://doi.org/10.1523/JNEUROSCI.2586-19.2020>
- Le, N. M., Yildirim, M., Wang, Y., Sugihara, H., Jazayeri, M., & Sur, M. (2022). *Mixture of Learning Strategies Underlies Rodent Behavior in Dynamic Foraging* [Preprint]. Neuroscience. <https://doi.org/10.1101/2022.03.14.484338>

- Lenow, J. K., Constantino, S. M., Daw, N. D., & Phelps, E. A. (2017). Chronic and Acute Stress Promote Overexploitation in Serial Decision Making. *The Journal of Neuroscience*, *37*(23), 5681–5689. <https://doi.org/10.1523/JNEUROSCI.3618-16.2017>
- Levy, A., Konidaris, G., Platt, R., & Saenko, K. (2017). *Learning Multi-Level Hierarchies with Hindsight*. <https://doi.org/10.48550/ARXIV.1712.00948>
- Li, F., Cao, W. Y., Li, M. B., Xu, Y., Zhang, J. W., Zhang, J. Y., Luo, X. G., Dai, R. P., Zhou, X. F., & Li, C. Q. (2012). A simple method for detection of food foraging behavior in the rat: Involvement of NMDA and dopamine receptors in the behavior. *Neuroscience*, *205*, 73–80. <https://doi.org/10.1016/j.neuroscience.2011.12.057>
- Li, F., Li, M., Cao, W., Xu, Y., Luo, Y., Zhong, X., Zhang, J., Dai, R., Zhou, X.-F., Li, Z., & Li, C. (2012). Anterior cingulate cortical lesion attenuates food foraging in rats. *Brain Research Bulletin*, *88*(6), 602–608. <https://doi.org/10.1016/j.brainresbull.2012.05.015>
- Li, S. W., Zeliger, O., Strahs, L., Báez-Mendoza, R., Johnson, L. M., McDonald Wojciechowski, A., & Williams, Z. M. (2022). Frontal neurons driving competitive behaviour and ecology of social groups. *Nature*, *603*(7902), 661–666. <https://doi.org/10.1038/s41586-021-04000-5>
- Lima, S. L. (1998). Stress and Decision Making under the Risk of Predation: Recent Developments from Behavioral, Reproductive, and Ecological Perspectives. In *Advances in the Study of Behavior* (Vol. 27, pp. 215–290). Elsevier. [https://doi.org/10.1016/S0065-3454\(08\)60366-6](https://doi.org/10.1016/S0065-3454(08)60366-6)
- March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.
- McNamara, J. M. (1982). Optimal patch use in a stochastic environment. *Theoretical Population Biology*, *21*(2), 269–288. [https://doi.org/10.1016/0040-5809\(82\)90018-1](https://doi.org/10.1016/0040-5809(82)90018-1)

- McNamara, J. M., & Houston, A. I. (1985). Optimal foraging and learning. *Journal of Theoretical Biology*, *117*(2), 231–249. [https://doi.org/10.1016/S0022-5193\(85\)80219-8](https://doi.org/10.1016/S0022-5193(85)80219-8)
- McNickle, G. G., & Cahill, J. F. (2009). Plant root growth and the marginal value theorem. *Proceedings of the National Academy of Sciences*, *106*(12), 4747–4751. <https://doi.org/10.1073/pnas.0807971106>
- Merkle, J. A., Fortin, D., & Morales, J. M. (2014). A memory-based foraging tactic reveals an adaptive mechanism for restricted space use. *Ecology Letters*, *17*(8), 924–931. <https://doi.org/10.1111/ele.12294>
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Henry Holt and Co. <https://doi.org/10.1037/10039-000>
- Miller, K. J., & Venditto, S. J. C. (2021). Multi-step planning in the brain. *Current Opinion in Behavioral Sciences*, *38*, 29–39. <https://doi.org/10.1016/j.cobeha.2020.07.003>
- Mobbs, D., Trimmer, P. C., Blumstein, D. T., & Dayan, P. (2018). Foraging for foundations in decision neuroscience: Insights from ethology. *Nature Reviews Neuroscience*, *19*(7), 419–427. <https://doi.org/10.1038/s41583-018-0010-7>
- Momennejad, I. (2020). Learning Structures: Predictive Representations, Replay, and Generalization. *Current Opinion in Behavioral Sciences*, *32*, 155–166. <https://doi.org/10.1016/j.cobeha.2020.02.017>
- Morimoto, J. (2019). Foraging decisions as multi-armed bandit problems: Applying reinforcement learning algorithms to foraging data. *Journal of Theoretical Biology*, *467*, 48–56. <https://doi.org/10.1016/j.jtbi.2019.02.002>
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520. <https://doi.org/10.1007/s00213-006-0502-4>

- Noel, J.-P., Caziot, B., Bruni, S., Fitzgerald, N. E., Avila, E., & Angelaki, D. E. (2021). Supporting generalization in non-human primate behavior by tapping into structural knowledge: Examples from sensorimotor mappings, inference, and decision-making. *Progress in Neurobiology*, *201*, 101996. <https://doi.org/10.1016/j.pneurobio.2021.101996>
- O'Fallon, S., Drager, K., Zhao, A., Suarez, A., & Pinter-Wollman, N. (2023). Foraging behaviour affects nest architecture in a cross-species comparison of ant nests. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1884), 20220146. <https://doi.org/10.1098/rstb.2022.0146>
- O'Reilly, R. C. (2010). The What and How of prefrontal cortical organization. *Trends in Neurosciences*, *33*(8), 355–361. <https://doi.org/10.1016/j.tins.2010.05.002>
- Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The Curse of Planning: Dissecting Multiple Reinforcement-Learning Systems by Taxing the Central Executive. *Psychological Science*, *24*(5), 751–761. <https://doi.org/10.1177/0956797612463080>
- Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946. <https://doi.org/10.1073/pnas.1312011110>
- Pateria, S., Subagdja, B., Tan, A., & Quek, C. (2022). Hierarchical Reinforcement Learning: A Comprehensive Survey. *ACM Computing Surveys*, *54*(5), 1–35. <https://doi.org/10.1145/3453160>
- Pearson, J. M., Watson, K. K., & Platt, M. L. (2014). Decision Making: The Neuroethological Turn. *Neuron*, *82*(5), 950–965. <https://doi.org/10.1016/j.neuron.2014.04.037>

- Peat, J., & Goulson, D. (2005). Effects of experience and weather on foraging rate and pollen versus nectar collection in the bumblebee, *Bombus terrestris*. *Behavioral Ecology and Sociobiology*, *58*(2), 152–156. <https://doi.org/10.1007/s00265-005-0916-8>
- Pereira-Obilinovic, U., Hou, H., Svoboda, K., & Wang, X.-J. (2022). *Brain mechanism of foraging: Reward-dependent synaptic plasticity or neural integration of values?* [Preprint]. bioRxiv. <https://doi.org/10.1101/2022.09.25.509030>
- Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford University Press.
- Pitkow, X., & Angelaki, D. E. (2017). Inference in the Brain: Statistics Flowing in Redundant Population Codes. *Neuron*, *94*(5), 943–953. <https://doi.org/10.1016/j.neuron.2017.05.028>
- Pomerleau, D. A. (1991). Efficient Training of Artificial Neural Networks for Autonomous Navigation. *Neural Computation*, *3*(1), 88–97. <https://doi.org/10.1162/neco.1991.3.1.88>
- Pouncy, T., Tsvividis, P., & Gershman, S. J. (2021). What Is the Model in Model-Based Planning? *Cognitive Science*, *45*(1). <https://doi.org/10.1111/cogs.12928>
- Pretelli, I., Ringen, E., & Lew-Levy, S. (2022). Foraging complexity and the evolution of childhood. *Science Advances*, *8*(41), eabn9889. <https://doi.org/10.1126/sciadv.abn9889>
- Radner, R., & Rothschild, M. (1975). On the allocation of effort. *Journal of Economic Theory*, *10*(3), 358–376.
- Ramakrishnan, A., Hayden, B. Y., & Platt, M. L. (2019). Local field potentials in dorsal anterior cingulate sulcus reflect rewards but not travel time costs during foraging. *Brain and Neuroscience Advances*, *3*, 239821281881793. <https://doi.org/10.1177/2398212818817932>

- Rathore, S. K., Das, G. K., Yadav, S. S., Yadav, S., & Yadav, A. S. (2022). Survey on Foraging Behavior Algorithms. *Proceedings of the 4th International Conference on Information Management & Machine Intelligence*, 1–6.
<https://doi.org/10.1145/3590837.3590890>
- Reid, A. K., & Staddon, J. E. R. (1998). A dynamic route finder for the cognitive map. *Psychological Review*, *105*(3), 585–601. <https://doi.org/10.1037/0033-295X.105.3.585>
- Rudebeck, P. H., & Izquierdo, A. (2021). Foraging with the frontal cortex: A cross-species evaluation of reward-guided behavior. *Neuropsychopharmacology*.
<https://doi.org/10.1038/s41386-021-01140-0>
- Rushworth, M. F. S., Kolling, N., Sallet, J., & Mars, R. B. (2012). Valuation and decision-making in frontal cortex: One or many serial or parallel systems? *Current Opinion in Neurobiology*, *22*(6), 946–955. <https://doi.org/10.1016/j.conb.2012.04.011>
- Rushworth, M. F. S., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron*, *70*(6), 1054–1069. <https://doi.org/10.1016/j.neuron.2011.05.014>
- Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic Drugs Modulate Learning Rates and Perseveration in Parkinson's Patients in a Dynamic Foraging Task. *Journal of Neuroscience*, *29*(48), 15104–15114. <https://doi.org/10.1523/JNEUROSCI.3524-09.2009>
- Scholl, J., Kolling, N., Nelissen, N., Wittmann, M. K., Harmer, C. J., & Rushworth, M. F. S. (2015). The Good, the Bad, and the Irrelevant: Neural Mechanisms of Learning Real and Hypothetical Rewards and Effort. *Journal of Neuroscience*, *35*(32), 11233–11251. <https://doi.org/10.1523/JNEUROSCI.0396-15.2015>

- Scholl, J., Trier, H. A., Rushworth, M. F. S., & Kolling, N. (2022). The effect of apathy and compulsivity on planning and stopping in sequential decision-making. *PLOS Biology*, 20(3), e3001566. <https://doi.org/10.1371/journal.pbio.3001566>
- Schultz, W. (2013). Updating dopamine reward signals. *Current Opinion in Neurobiology*, 23(2), 229–238. <https://doi.org/10.1016/j.conb.2012.11.012>
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, 275(5306), 1593–1599. <https://doi.org/10.1126/science.275.5306.1593>
- Schultz, W., & Dickinson, A. (2000). Neuronal Coding of Prediction Errors. *Annual Review of Neuroscience*, 23(1), 473–500. <https://doi.org/10.1146/annurev.neuro.23.1.473>
- Seamans, J. K., Floresco, S. B., & Phillips, A. G. (1995). Functional differences between the prelimbic and anterior cingulate regions of the rat prefrontal cortex. *Behavioral Neuroscience*, 109(6), 1063–1073. <https://doi.org/10.1037/0735-7044.109.6.1063>
- Shahidi, N., Parajuli, A., Franch, M., Schrater, P., Wright, A., Pitkow, X., & Dragoi, V. (2019). *Population coding of strategic variables during foraging in freely-moving macaques* [Preprint]. Neuroscience. <https://doi.org/10.1101/811992>
- Shahnazian, D., & Holroyd, C. B. (2018). Distributed representations of action sequences in anterior cingulate cortex: A recurrent neural network approach. *Psychonomic Bulletin & Review*, 25(1), 302–321. <https://doi.org/10.3758/s13423-017-1280-1>
- Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, 19(10), 1286–1291. <https://doi.org/10.1038/nn.4384>
- Shenhav, A., Straccia, M. A., Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience*, 17(9), 1249–1254. <https://doi.org/10.1038/nn.3771>

- Shteingart, H., & Loewenstein, Y. (2014). Reinforcement learning and human behavior. *Current Opinion in Neurobiology*, 25, 93–98.
<https://doi.org/10.1016/j.conb.2013.12.004>
- Silston, B., Wise, T., Qi, S., Sui, X., Dayan, P., & Mobbs, D. (2021). Neural encoding of perceived patch value during competitive and hazardous virtual foraging. *Nature Communications*, 12(1), 5478. <https://doi.org/10.1038/s41467-021-25816-9>
- Singh, S., Barto, A. G., & Chentanez, N. (2004). Intrinsically Motivated Reinforcement Learning: *Proceedings of the 2004 Conference*. Advances in Neural Information Processing Systems 17, Fort Belvoir, VA. <https://doi.org/10.21236/ADA440280>
- Skinner, B. F. (1965). *Science and human behavior* (First Free Press Paperback edition). The Free Press.
- Slotow, R., & Coumi, N. (2000). VIGILANCE IN BRONZE MANNIKIN GROUPS: THE CONTRIBUTIONS OF PREDATION RISK AND INTRA-GROUP COMPETITION. *Behaviour*, 137(5), 565–578.
<https://doi.org/10.1163/156853900502213>
- Smith, E. A., Bettinger, R. L., Bishop, C. A., Blundell, V., Cashdan, E., Casimir, M. J., Christenson, A. L., Cox, B., Dyson-Hudson, R., Hayden, B., Richerson, P. J., Roth, E. A., Simms, S. R., & Stini, W. A. (1983). Anthropological Applications of Optimal Foraging Theory: A Critical Review [and Comments and Reply]. *Current Anthropology*, 24(5), 625–651. JSTOR.
- Song, H. F., Yang, G. R., & Wang, X.-J. (2016). *Reward-based training of recurrent neural networks for cognitive and value-based tasks* [Preprint]. Neuroscience.
<https://doi.org/10.1101/070375>
- Stephens, D. W., Brown, J. S., & Ydenberg, R. C. (Eds.). (2007). *Foraging: Behavior and ecology*. Univ. of Chicago Press.

- Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton University Press.
- Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching Behavior and the Representation of Value in the Parietal Cortex. *Science*, *304*(5678), 1782–1787.
<https://doi.org/10.1126/science.1094765>
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.
- Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, *112*(1–2), 181–211. [https://doi.org/10.1016/S0004-3702\(99\)00052-1](https://doi.org/10.1016/S0004-3702(99)00052-1)
- Taren, A. A., Venkatraman, V., & Huettel, S. A. (2011). A Parallel Functional Topography between Medial and Lateral Prefrontal Cortex: Evidence and Implications for Cognitive Control. *The Journal of Neuroscience*, *31*(13), 5026–5031.
<https://doi.org/10.1523/JNEUROSCI.5762-10.2011>
- Tervo, D. G. R., Kuleshova, E., Manakov, M., Proskurin, M., Karlsson, M., Lustig, A., Behnam, R., & Karpova, A. Y. (2021). The anterior cingulate cortex directs exploration of alternative strategies. *Neuron*, *109*(11), 1876-1887.e6.
<https://doi.org/10.1016/j.neuron.2021.03.028>
- Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, *2*(4), i–109. <https://doi.org/10.1037/h0092987>
- van Heukelum, S., Mars, R. B., Guthrie, M., Buitelaar, J. K., Beckmann, C. F., Tiesinga, P. H. E., Vogt, B. A., Glennon, J. C., & Havenith, M. N. (2020). Where is Cingulate Cortex? A Cross-Species View. *Trends in Neurosciences*, *43*(5), 285–299.
<https://doi.org/10.1016/j.tins.2020.03.007>

- Verguts, T., Vassena, E., & Silvetti, M. (2015). Adaptive effort investment in cognitive and physical tasks: A neurocomputational model. *Frontiers in Behavioral Neuroscience*, 9. <https://doi.org/10.3389/fnbeh.2015.00057>
- Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., Oude Lohuis, M. N., & Mainen, Z. F. (2020). Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas. *Neuron*, 106(1), 166-176.e6. <https://doi.org/10.1016/j.neuron.2020.01.017>
- Wells, V., Carrigan, M., & Athwal, N. (2023). Pandemic-driven consumer behaviour: A foraging exploration. *Marketing Theory*, 147059312311756. <https://doi.org/10.1177/14705931231175695>
- Wikenheiser, A. M., Stephens, D. W., & Redish, A. D. (2013). Subjective costs drive overly patient foraging strategies in rats on an intertemporal foraging task. *Proceedings of the National Academy of Sciences*, 110(20), 8308–8313. <https://doi.org/10.1073/pnas.1220738110>
- Winfield, A. F. (2009). Foraging Robots. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 3682–3700). Springer New York. https://doi.org/10.1007/978-0-387-30440-3_217
- Wise, T., Emery, K., & Radulescu, A. (2023). Naturalistic reinforcement learning. *Trends in Cognitive Sciences*, S1364661323002127. <https://doi.org/10.1016/j.tics.2023.08.016>
- Wispiński, N., Butcher, A., Mathewson, K. W., Chapman, C. S., Botvinick, M. M., & Pilarski, P. M. (2023). Adaptive patch foraging in deep reinforcement learning agents. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=a0T3nOP9sB>
- Wittmann, M. K., Kolling, N., Akaishi, R., Chau, B. K. H., Brown, J. W., Nelissen, N., & Rushworth, M. F. S. (2016). Predictive decision making driven by multiple time-

- linked reward representations in the anterior cingulate cortex. *Nature Communications*, 7(1), 12327. <https://doi.org/10.1038/ncomms12327>
- Wolfe, J. M. (2013). When is it time to move to the next raspberry bush? Foraging rules in human visual search. *Journal of Vision*, 13(3), 10–10. <https://doi.org/10.1167/13.3.10>
- Wu, J., & Aberer, K. (2003). Swarm Intelligent Surfing in the Web. In J. M. C. Lovelle, B. M. G. Rodríguez, J. E. L. Gayo, M. Del Puerto Paule Ruiz, & L. J. Aguilar (Eds.), *ICWE 2003: Web Engineering* (Vol. 2722, pp. 431–440). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45068-8_80
- Wu, Z., Kwon, M., Daptardar, S., Schrater, P., & Pitkow, X. (2020). Rational thoughts in neural codes. *Proceedings of the National Academy of Sciences*, 117(47), 29311–29320. <https://doi.org/10.1073/pnas.1912336117>
- Yoo, S. B. M., Tu, J. C., Piantadosi, S. T., & Hayden, B. Y. (2020). The neural basis of predictive pursuit. *Nature Neuroscience*, 23(2), 252–259. <https://doi.org/10.1038/s41593-019-0561-6>
- Zhang, J., Gong, X., Fougny, D., & Wolfe, J. M. (2015). Using the past to anticipate the future in human foraging behavior. *Vision Research*, 111, 66–74. <https://doi.org/10.1016/j.visres.2015.04.003>





MVT	MF-RL	MB-RL	MB-HRL
			
leave a patch when the latest reward is lower than the average reward rate of the environment.	cache reward and state values, grant preference to the choice (stay vs. leave) with higher value.	an internal model of the environment allows for planning lines of behavior (harvest/leave) and their probable outcomes.	contiguous actions are grouped into options, the average reward value is monitored across distinct levels of abstraction, in order to dictate when to leave or persist with goals/subgoals.
Retrospective		Prospective	

Figure 1. Patch-leaving decision criteria according to the MVT and RL models. MVT compares the return of the current patch to the global environment average return. MF-RL favors the choice (staying or leaving) with higher value. Model-based RL models use an internal model to predict long-term consequences of available choices, favoring the decision that maximizes the likelihood of reward maximization. Model-based models can describe foraging behavior more accurately but are more complex. A hierarchical organization can simplify these internal representations by modularizing the foraging task in hierarchical subcomponents. This in turn captures features of behavior better than its non-hierarchical analogs. MVT: Marginal Value Theorem; MF-RL: model-free reinforcement learning, MB-RL: model-based reinforcement learning; MB-HRL: model-based hierarchical reinforcement learning.

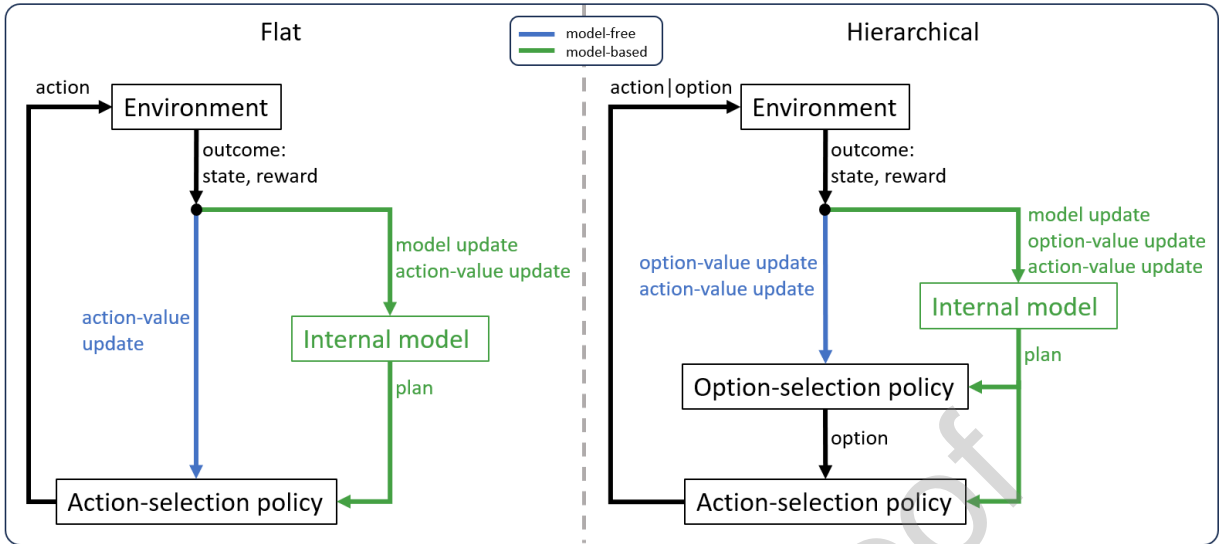


Figure 2. Feature comparison of MF-RL (blue connections) and MB-RL (green connections), for both flat and hierarchical versions. In flat RL there is no concept of sub-goal or secondary reward attainment. MF-RL caches reward values updated via RPEs in order to drive learning retrospectively, while flat MB-RL uses an internal model of the world to simulate possible paths and prospectively guide (plan) decisions. MF-HRL is supported by RPEs at different levels that drive learning of action values at the level of actions and policy values at the level of options (subgoals). MB-HRL leverages the advantages of options for goal-directed planning, which can serve well for complex, stochastic and/or non-stationary environments.

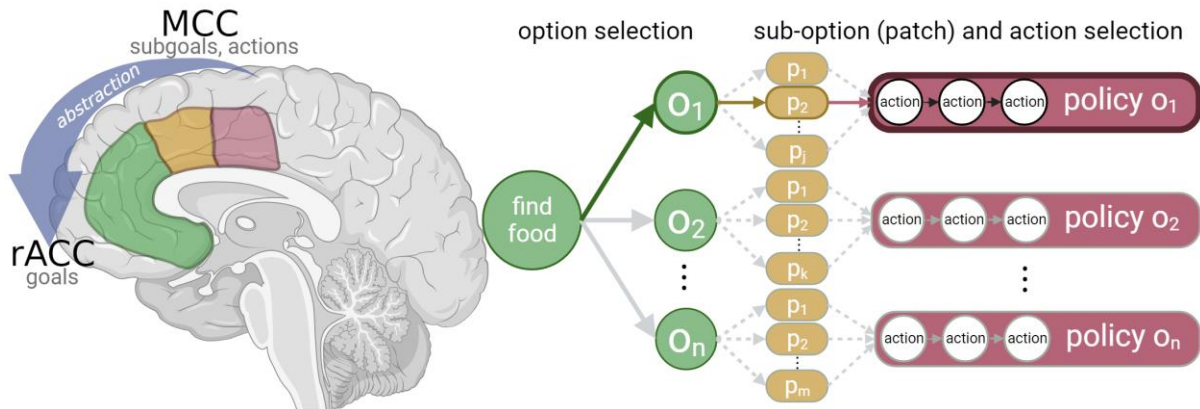


Figure 3. Foraging as a hierarchical task and its cingulate cortex correlates. Rostral ACC (rACC), including neighboring medial orbitofrontal cortex (mOFC) regions (green shade) implement high-level task goals and action schemas (such as *find food*) by selecting and maintaining high-level options (O_1, O_2, \dots, O_n , such as *hunting, fishing*, etc.). Anterior midcingulate cortex (aMCC; yellow shade) selects and maintains sub-options (p_1, p_2, \dots, p_m , such as *patch 1, patch 2*, etc.) that accomplish less abstract lower-level sub-goals, and posterior midcingulate cortex (pMCC) including neighboring pre-supplementary motor cortex (pre-SMA) areas (red shade) regulate the individual low-level action sequences driven by option-specific policies. At any given time, the ACC selects or maintains an option (green arrow), which itself selects or maintains engagement with a patch (brown arrow), which in turn invokes a harvesting-related action policy (red arrow). When the rewards received in a patch are worse than expected, the option level increases control over the patch level, eventually driving a switch to a different patch (despite travelling costs) such as from one tree to another. When the rewards received across the entire task are worse than expected, the task level increases control over the option level, eventually driving a switch to a different option (despite the travelling costs) such as from hunting to fishing. Control is then released when the received outcomes (at each level) are better than expected. See van Heukelum et al. (2020) for anatomical definitions; for clarity, we reserve the terms ACC for the entire region described (rACC, aMCC, pMCC), and *rACC* for what Heukelum et al (2020) call ACC. Created with BioRender.

Highlights

- Foraging behavior can be difficult to explain in complex stochastic environments.
- The role of the anterior cingulate cortex during foraging is uncertain and disputed.
- Model-free RL approaches neglect the use of environmental information that animals seem to leverage.
- Model-based RL approaches are limited by the processing demands of online planning during foraging.
- The hierarchical structure of foraging problems and a hierarchical reinforcement learning mechanism implemented by ACC can provide insight into the neural computations responsible for foraging behavior.