How multiple learning systems contribute to naturalistic patch foraging

John P. Kasarda[1], **Jyotika Bahuguna**[2], **Angela Zhang**[3], **Hua Tong**[3], **Yuan Tan**[3], **Ruizi Wang**[3], **Michael J. Tarr**[1,4,*], **and Timothy Verstynen**[1,4,*]

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh, 15213, USA
[2]Laboratoire de Neurosciences Cognitives et Adaptatives (LNCA), Faculté de Psychologie, Université de Strasbourg, Strasbourg, France
[3]Entertainment Technology Center, Carnegie Mellon University,Pittsburgh, 15213, USA
[4]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, 15213, USA
[*]corresponding author(s): Timothy Verstynen (timothyv@andrew.cmu.edu), Michael J Tarr (michaeltarr@cmu.edu)

ABSTRACT

Optimal patch foraging requires the ability to discern valuable resources and harvest them efficiently. This process requires integrating across multiple learning systems, including perceptual, structural, and value-based mechanisms. We explored how these varied learning systems contribute to foraging efficiency through a novel virtual foraging videogame. Human participants (N=60) navigated a virtual world with four neighborhoods (patches) populated with houses, searching for "bitter Oomplets," characterized by the union of specific body colors and textures. Each neighborhood had a different overall probability of containing bitter Oomplets and, within each, specific house locations predicted the relative rate of bitter Oomplets. Participants proficiently learned to discriminate Oomplet types and strategically allocate their time to higher-probability neighborhoods. However, they did not learn the spatial predictors within neighborhoods. Importantly, we found no interaction between these two distinct learning abilities in determining foraging ability, suggesting that the impact of the perceptual and reward learning systems are independent of one another in human foraging efficiency.

Main

Patch foraging is one of the most ubiquitous behaviors in nature[1–3]. Consider, for example, a lemur searching for ripe fruit in a forest. She must first decide which tree to climb and search, with each serving as a potential resource patch. Once in the tree, she needs to discriminate ripe fruit from unripe fruit by its appearance and infer where clusters of ripe fruit are most likely to be found. Eventually, she faces a decision: continue harvesting from the current tree or move to a new one in search of a potentially more abundant yield.

This scenario represents a *nested patch foraging problem*. More generally, patch foraging is a distinct type of decision-making challenge that can be represented as an acyclic graph of choices[4], including how to search for patches, whether to accept a discovered patch for harvesting, and when to stop harvesting and move on. In the nested version of patch foraging, the harvesting phase itself contains another acyclic graph of *explore → patch select → harvest* decisions. While the ecological literature has typically examined these decisions only in terms of the overt behavior, such as the optimality of Lévy flights during search[5] or the optimal timing to stop harvesting[6], cognitive science and related fields have focused more on the learning mechanisms that support the foraging and decision-making process. This research often treats patch foraging as a variant of the standard reinforcement learning problem[7–9], particularly one that involves managing the speed-accuracy tradeoff during decision-making[10]. This value learning framework has been used to explain foraging patterns across a wide range of species, including bees[11], bats[12], humans[10,13,14], and machines[15,16].

Beyond reinforcement learning, naturalistic foraging scenarios, like the one faced by our hypothetical lemur, also interact with other learning systems. To forage efficiently, she must learn to make accurate perceptual judgments about the ripeness of the fruit[17] and develop knowledge of the internal structure of the trees[18] in order to identify the best locations for harvesting. How well she learns these skills is likely have a substantial impact on the speed and accuracy of harvesting decisions and, in turn, on overall patch foraging efficiency. Perceptual foraging paradigms suggest that value learning and perceptual expertise may interact with each other[19]; however, in the typical experimental paradigm, the perceptual discrimination and foraging actions exist in the same space (i.e., visual representations). As such, we still do not fully understand how these other learning interact with value learning to shape optimal foraging behavior in traditional patch foraging tasks.

In this study, we examine how learning about the environment supports and shapes foraging decisions in a naturalistic foraging paradigm. We use a novel video game, "BitterBuster", in which participants navigate a virtual candy-themed world in search of candy creatures known as "Oomplets". Oomplets are distinguished as bitter or sweet by perceptual features that participants must learn through trial and error[20]. Players explore four neighborhoods (corresponding to foraging patches) containing various houses, visiting each house individually to collect bags of candy and search for bitter Oomplets. Each neighborhood has a different overall probability of containing bitter Oomplets, and within each neighborhood, the likelihood of encountering bitter Oomplets depends on the location of a house within the neighborhood. By analyzing behavior in these exploration and selection tasks, we are better able to understand how environmental factors influence foraging proficiency and to identify the neural systems that guide naturalistic foraging decisions.
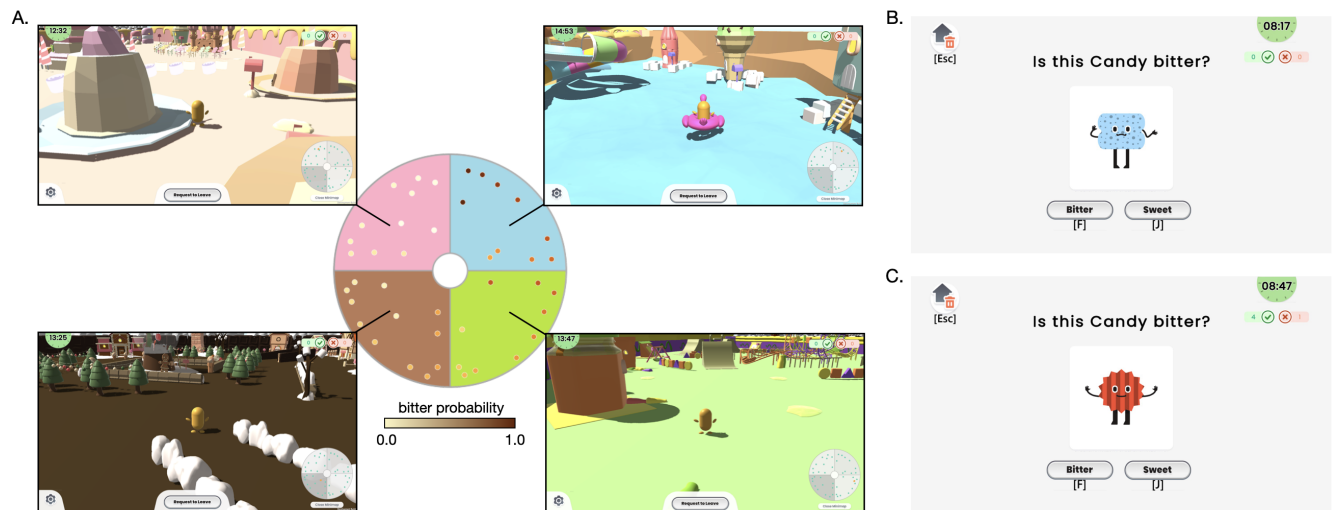
Results

Figure 1. (A) Top down view of the Oomplet kingdom, and screenshots of its four neighborhoods. (B) A bitter Oomplet in the selection task window. (C) A sweet Oomplet in the selection task window.

To play BitterBuster, participants are instructed to find and correctly identify as many bitter Oomplets as possible within the time allotted for each round of testing. To accomplish this, participants explore four neighborhoods within a virtual candy world (Figure 1A). Starting in the central arena, participants select a neighborhood (patch selection) and approach one of the houses scattered around the area (represented by colored circles in the inset of Figure 1A). A house can be explored by entering it, at which point the participant may begin harvesting Oomplets. Upon entering, the screen displays a new window where participants are serially presented with images of Oomplets that may be either bitter (Figure 1B) or sweet (Figure 1C). Participants make a binary classification based on the perceptual features of each Oomplet and are then given feedback on whether their selection was correct or not.

Participants can choose to exit the house at any point or can continue classifying Oomplets until all Oomplets inside a given house have been categorized. Participants are not told how many Oomplets are in a house, with the number Oomplets per house ranging from 8-12. After exiting, the view returns to the neighborhood screen, where participants can either continue to another house or leave the neighborhood altogether. A selection round ends when either all available Oomplets have been harvested across all neighborhoods or the time limit has been reached. At the end of each round, participants receive feedback indicating the proportion of bitter Oomplets they correctly identified (hit rate) and proportion of sweet Oomplets incorrectly identified as bitter (false alarms). On the next round all houses are repopulated with Oomplets using the same sampling probability for all testing rounds. Participants in our sample (N=60) were tested across three rounds of varying lengths (20 minutes, 10 minutes, 10 minutes). Participants were informed of the round durations before testing. The first round was afforded more time in order to give participants sufficient opportunity to explore the virtual world as thoroughly as possible.

For the purposes of our analyses each visit to a house was considered a trial and each *correctly* classified bitter Oomplet was considered a reward. Looking at the average rewards collected, relative to the time when a neighborhood was abandoned, we see two relevant patterns (Figure 2A). First, the overall average rewards collected increases with training, regardless of the trial relative to patch abandonment. Second, the average number of collected rewards (black line in Figure 2A) shows a
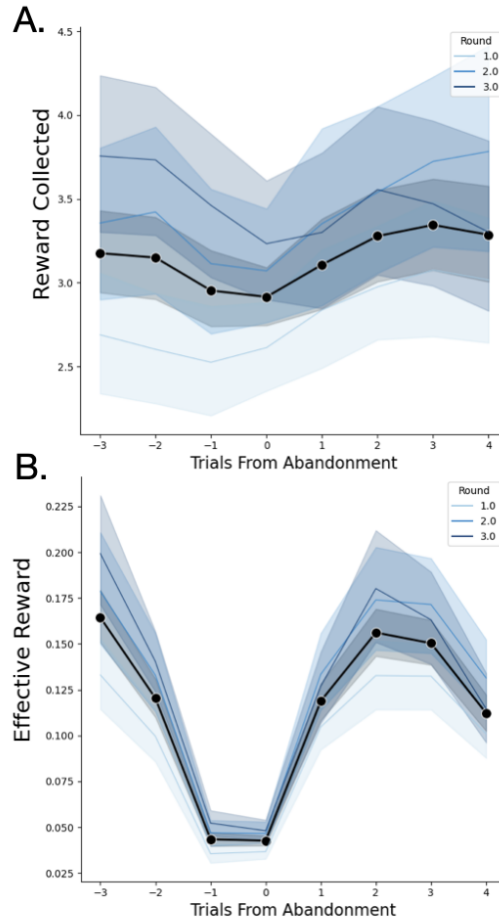
Figure 2. Effective foraging rate. (A) Collected rewards – correctly classified bitter Oomplets – on each house visit time locked to the decision to leave a neighborhood (abandonment). Color hue indicates round and black line indicates average performance across rounds. (B) Expected reward rate – the rate of correctly classified bitter Oomplets relative to distance between houses – time-locked to patch abandonment. Same color coding scheme as in panel A.

slight dip around the time of patch abandonment, but the magnitude of this drop is not particularly large. This reflects a drop in harvesting rate as participants make their way through a patch.

According to the Marginal Value Theorem[6] (MVT), optimal foraging necessitates balancing rate of reward collection against expected effort to travel to new resource patches. To capture this balance, we created a measure called effective reward rate (see *Methods*). This reflects the immediate return on rewards (i.e., correctly identified bitter Oomplets), normalized by the distance between houses, approximating the effort to transit to another sample of Oomplets. As participants approached the decision to leave a neighborhood (or patch), the effective reward rate declined, and then increased again after switching to a new neighborhood (Figure 2B). The drop in the effective reward rate near patch abandonment was significantly higher than the drop in the overall collected reward, indicating that participants were tracking the effort needed to get to a new patch of bitter Oomplets when making their harvesting decisions.

When we compare the effective reward rate at the trial just before a neighborhood abandonment, against the average reward collected on that trial, we see that this non-monotonic relationship saturates at higher rates of effective rewards (Figure 3A-C). This relationship also strengthens with training, revealing two important patterns. First, this shows how harvesting becomes less efficient at the end of a patch harvest. Both premature exit of the patch (effective rewards of $0.00 - 0.02$), as well as lingering on the patch too long (effective reward > 0.08), both lead to lower collected rewards. Second, this bottleneck on harvesting efficiency gets stronger with learning, suggesting that participants are learning the return rate for each patch.

The results shown in Figure 2 and Figure 3 indicate that the effective reward rate increases with practice. To quantify this effect, we looked at the overall average effective reward rate across rounds. As training progressed, the average effective reward rate increased from a mean of 0.078 in the first round to 0.117 in the final round (Figure 4A). This effect was reliable across
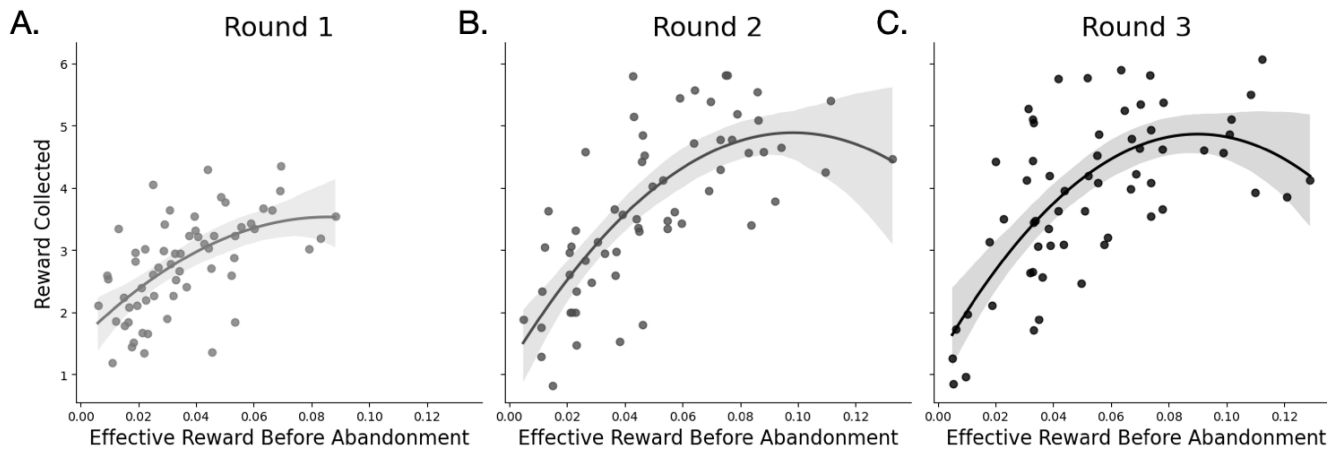
Figure 3. Effective foraging rate. The relationship between average collected rewards and average effective reward rate for the first (A), second (B), and third (C) round of training. Each point represents the average effective reward before patch abandonment and the total average reward acquired per participant in a given round.

participants as determined as a main effect of round in a linear, mixed-effects regression model ($\beta = 0.02, SE = 0.003, p < 0.001$ Supplementary Table 1). This confirms that, with three rounds of practice, participants were able to learn to forage effectively in our naturalistic foraging task.

Subsequent to confirming that human participants can learn to forage effectively in the BitterBuster environment, we next examined perceptual, value, and structure learning during the foraging task. To assess perceptual sensitivity over time, we calculated the d' for Oomplet classification (see *Methods*). Perhaps unsurprisingly, d' values increased with practice, beginning with an average of 1.55 in Round 1 and rising to 3.01 in Round 3 ($\beta = 0.72$, $SE = 0.07$, $p < 0.001$ Supplementary Table 2; Figure 4B). These results indicate that participants were able to learn the perceptual features that distinguish bitter from sweet Oomplets, thereby improving their classification ability over time (see also[20]).

To estimate value learning, we calculated a neighborhood bias score (see *Methods*) reflecting the proportion of time spent in each neighborhood weighted by that neighborhood's overall bitter Oomplet probability. This bias score ranges from 0 to 1, with higher values indicating that a participant spent more time in neighborhoods with higher probabilities of bitter Oomplets. On average, neighborhood bias increased with practice, rising from 0.50 in Round 1 to 0.62 in Round 3 ($\beta = 0.23$, $SE = 0.01$, $p < 0.001$ Supplementary Table 3; Figure 4C). These results indicate that participants successfully learned to identify and value neighborhoods with higher likelihoods of having bitter Oomplets.

Finally, at a more granular level, we examined whether participants could learn the within-neighborhood structural patterns that predicted bitter Oomplet probability. To assess this, we created a house bias score. This was calculated by ranking each house within a neighborhood according to its actual bitter probability and comparing that to the participant's search order using a Spearman's rank-order correlation coefficient to quantify their alignment (see *Methods*). In contrast to neighborhood bias, house bias did not show significant improvement with practice. The mean value was maintained at 0.66 through Rounds 1 to 3 ($\beta = 0.00$, $SE = 0.01$, $p = 0.684$ Supplementary Table 4; Figure 4D). These results indicate that while participants were able to detect global patterns across neighborhoods, the finer-grained spatial patterns within neighborhoods were not reliably learned during the period of time allotted in our study (a single day of training on the BitterBuster task).

To summarize our results to this point, we show that overall foraging efficiency in the BitterBuster task improves with practice. This improvement coincides with gains in perceptual discrimination and patch value learning, but not with learning of within-patch latent structure. While providing some insight into human patch foraging behavior, we are primarily interested in understanding how disparate learning abilities contribute to patch foraging effectiveness. To this end, we examined how effective reward rate relates to perceptual and value learning – the two cognitive abilities that exhibited practice-related improvements. Figure 5 illustrates the relationships between d', neighborhood bias, and effective reward rate across all three training rounds. These plots reveal a clear emergence of correlations between d' and neighborhood bias in their contribution to effective reward, from early in training (Round 1) to later in training (Round 3). This trend suggests that both perceptual and value learning may independently relate to foraging ability, but may also interact with one another.

In order to formally test this possibility, we conducted a linear mixed-effects regression analysis to assess the influence of d' and neighborhood bias on effective reward (see *Methods*). The results showed a significant main effect of d' ($\beta = 0.01$, $SE = 0.00$, $p = 0.001$) and a strong main effect of neighborhood bias ($\beta = 0.20$, $SE = 0.02$, $p < 0.001$; Supplementary
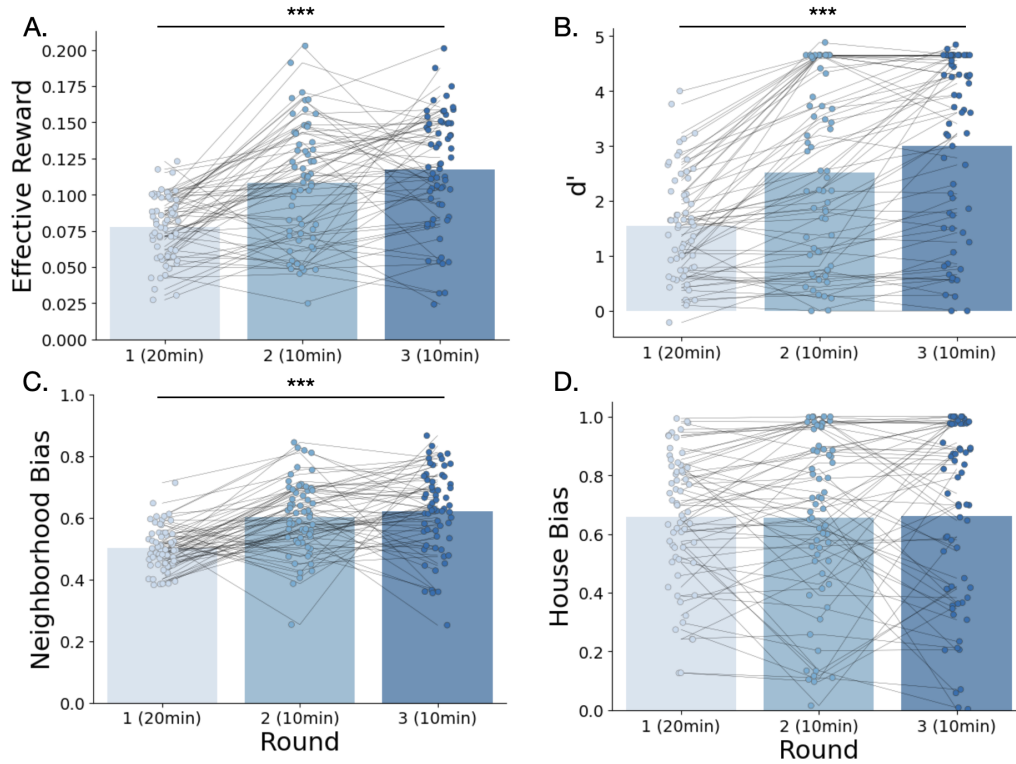
Figure 4. Performance measure means and distributions for all participants across all three rounds. Measures include: (A) Overall performance measured by Effective Reward, (B) perceptual discrimination measured by d', (C) patch value learning measured by Neighborhood Bias, and (D) patch structure learning measured by House Bias. *** $p < .001$

Table 5). However, we did not observe a significant interaction between these two factors ($\beta = -0.003$, $SE = 0.00$, $p = 0.68$). This indicates that overall foraging effectiveness (i.e., effective reward) in the BitterBuster task most likely reflects the additive contributions of perceptual and value learning independently.

Discussion

To understand the relative contributions of perceptual, value, and latent structure learning to human foraging efficiency, we developed a novel patch foraging task within a virtual environment. Our findings indicate that, even within a relatively short period of time, humans effectively learn the perceptual features that reliably discriminate viable resources (bitter Oomplets) and how to predict the relative value (probability of bitter Oomplets) of different patches. Interestingly, we did not observe reliable structure learning (house location predicting bitter Oomplet probability) within the allotted testing time, suggesting that if this ability manifests during foraging it requires more extensive experience than provided in our study. Most importantly, with respect to our primary hypothesis, we found that perceptual and value learning abilities function independently in enabling foraging effectiveness, suggesting that the cognitive processes contributing to patch foraging do not interact with one another.

Participant choices in our task mirrored the feedback-driven strategies expected in patch-foraging contexts[1,6,21–24]. While this does establish the ecological validity of foraging behavior in our virtual environment, one might also be concerned that value learning here is governed by a generic reinforcement-learning (RL) rule. However, prior work paints a more nuanced picture. For example, Constantino and Daw (2015)[14] found that a policy incorporating the opportunity costs emphasized by the MVT explained human foraging choices better than a standard RL model. Echoing this result, we observed a sharp decline in participants' effective reward rate (current returns + anticipated travel-time costs) as patch residence time increased, even though cumulative rewards continued to rise. This divergence supports the idea that feedback-driven learning during foraging integrates explicit cost terms into its value computations, rather than relying on a purely reward-based policy.

In our view, the most surprising outcome of our study is the apparent independence of perceptual and value learning in predicting patch foraging performance. At first glance, this seems to contradict a large body of work showing a tight coupling between these processes in perceptual-foraging settings. Bella-Fernández et al. (2022) review tasks in which observers search
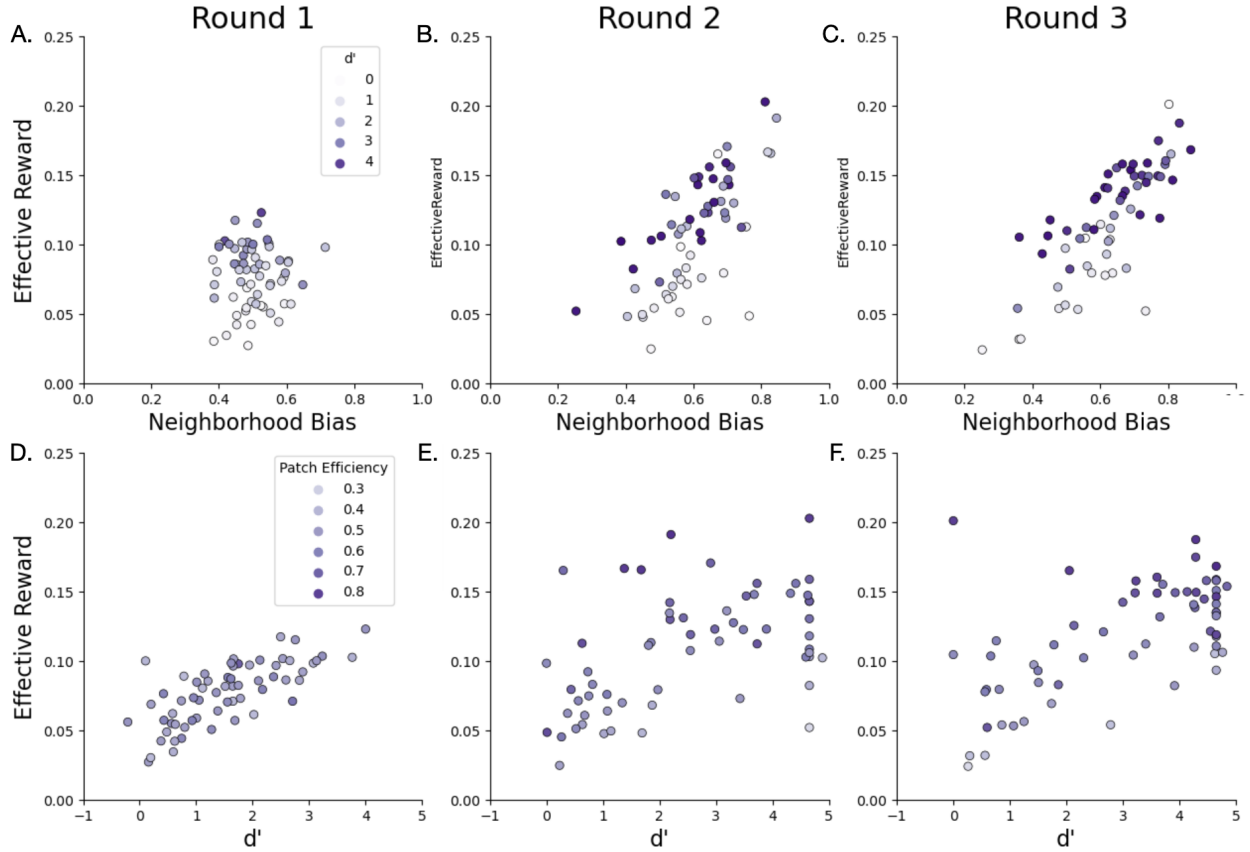
Figure 5. Effective Reward plotted against task performance measures for each training round. (A-C). Exploration performance - Neighborhood Bias - for Rounds 1, 2, and 3. (D-F). Selection performance - d' - for Rounds 1, 2, and 3. Color hue represents the performance measure value from the complementary task.

visually presented clustered targets and decide when to abandon a patch (i.e., look away). In these sorts of visual foraging tasks, sharper perceptual discrimination narrows uncertainty in value estimates, leading to better patch-leaving thresholds[19]. However, Wolfe and colleagues (2018) report the reverse: in a hybrid foraging task, learned target values steer visual attention, accelerating both search and departure decisions[25]. Together these studies suggest a reciprocal facilitation between perception and valuation. In this context, why did we not observe such an interaction? One possibility is that the Oomplet classification task is perceptually trivial with a reward mapping that is explicit and stationary. Given such low uncertainty in the Oomplet classification subtask, joint updating provides a minimal advantage. More specifically, keeping perceptual and value learning independent may be computationally less expensive under such low-complexity conditions. That is, the independence of patch and target features allows for rapid, independent convergence in each learning domain separately.

Another critical distinction between perceptual foraging and the type of foraging that we studied here may be in how value is represented. In visual foraging tasks, the reward is intrinsic and yoked with the process of visual recognition itself. Indeed, these sorts of tasks are more akin to basic visual search problems than more traditional foraging paradigms. This is *very* different from foraging tasks studied in ecological contexts (see[1,4,26] for review), a context which BitterBuster was designed to more closely emulate. In these scenarios the perceptual subtask, identifying viable targets to harvest, is more cognitively distinct, in both representation and implementation, from the exploration subtask of physically moving between patches. The rewards are also presented as more tangible points, rather than intrinsic recognition values. These differences may explain why perceptual learning and value learning were not found to interact in contributing to foraging effectiveness within our study.

Another area where our results diverge from prior literature is our failure to observe spatial structure learning within patches. Latent learning in spatial navigation involves forming internal cognitive maps without explicit instruction, but not necessarily without motivation. While Tolman originally argued that such learning occurs without rewards[18], more recent research suggests that reward signals shape and prioritize what is encoded. For instance, Doeller et al. (2008) found that the hippocampus

integrates spatial cues, but reward-related regions such as the striatum also play a role in structuring that information[27]. Likewise, a review by Gershman and Daw (2017) observed that dopaminergic signals, often tied to reward prediction, can enhance memory consolidation and structure learning in spatial tasks[28]. Schuck et al. (2016) demonstrated that even abstract state representations in the orbitofrontal cortex reflect latent structures that align with learned reward contingencies[29]. Thus, while latent structure learning can begin without feedback, rewards help to reinforce and stabilize the emerging cognitive maps. However, a single session may not provide enough reinforcement opportunities for these systems to prioritize and consolidate meaningful structure (e.g.,[30]), which could explain why we were unable to detect latent structure learning in our current study.

Future research on this topic should note a key shortcoming in our experiment's design. Although our findings were robust, we observed a strong average signal bias that emerged prior to the completion of the first round. The average d' scores increased significantly during the next two rounds, but the practical increase in the bias strength represented is marginal. Perhaps changing the Oomplet characteristics used in the perceptual discrimination task to be more difficult to detect (see[20]) could extend the learning window to match that of spatial navigation. We leave this to future studies.

In summary, we show that in a modular learning architecture for ecologically-valid foraging humans rapidly acquire perceptual cues and cost-sensitive value policies, but these streams appear to operate in parallel rather than an integrated fashion. We also observe that, within Bitterbuster, a single session potentially provides too little reinforcement for latent spatial maps to fully crystallize. This dissociation contrasts with classic visual-search foraging paradigms, suggesting that when perceptual classification is relatively straightforward and travel incurs meaningful costs, integrating perception and valuation yields no additional advantage. By identifying when learning systems act cooperatively versus independently, our findings refine theoretical boundaries of optimal foraging theory[6], underscoring the role of opportunity-cost signals in guiding patch-leaving decisions.

Methods

Participants

A total of 62 participants were initially recruited from the Carnegie Mellon University Psychology Department SONA System and compensated at a rate of 1 credit per hour. Data from two participants were excluded from the analysis due to software misconfiguration or failure to perform the task, leading to a final sample of 60 participants (23 male, mean age = 18.70 years). All procedures were approved and monitored by the local Institutional Review Board at Carnegie Mellon University.

Experimental design

At the start of each round, participants began in a small circular room in the middle of a candy kingdom with four doors leading to separate neighborhoods. Each neighborhood contained nine to twelve houses inhabited by candy-like creatures called Oomplets. Oomplets are configurable humanoid characters that can be optimized for complex perceptual discrimination tasks[20]. The task was to find and correctly identify as many bitter Oomplets as possible, based solely on their perceptual features. Participants were not instructed on how to discriminate between bitter and sweet Oomplets, using only in-game feedback learning to guide their decisions. Each neighborhood had its own maximum likelihood of having bitter Oomplets: 1.0, 0.75, 0.50, and 0.25. In addition, within each neighborhood, each house had a different probability of having bitter Oomplets. The house probability was multiplied by the neighborhood probability to determine the likelihood of bitter Oomplets at each house.

Within a neighborhood, the probability of bitter Oomplets at a house reduced in a clockwise direction relative to the entry point of that neighborhood (Figure 1A). The color of each house's mailbox was a cue as to the relative percentage of bitter Oomplets present in the house. Each mailbox used the same 3D model, but its hue was changed along a blue-to-orange spectrum, corresponding to the portion of bitter Oomplets in that house. Bluer hues reflected higher bitter Oomplet probability such that a deep blue/purple mailbox would correspond to a patch with mostly bitter Oomplets, a pink mailbox to an even distribution, and a light orange mailbox to mostly sweet Oomplets. Participants were not explicitly informed about either the mailbox color or clockwise probability cues – thus, exploration was necessary to learn their significance.

The exploration of the candy kingdom was periodically interrupted by the selection task. This task began when participants entered a house by pressing the *e* key when they were in close proximity to the door. During this time, the screen would switch to a 2D display (Figure 1B&C) where Oomplets were presented one at a time and participants were asked to classify them by pressing the *f* key for bitter and *j* key for sweet. Immediately after each selection participants were given visual feedback on the accuracy of their selection. An Oomplet's category was determined by two of ten possible attributes: hue and pattern. Bitter Oomplets were those which had a bluish hue and a spotted background[20]. All Oomplets with other colors and/or backgrounds were classified as sweet. When a participant correctly identified a bitter Oomplet, a reward sound played to accompany the visual feedback icon. In addition, an accuracy counter was updated in the upper right part of the screen. A total of eight to twelve Oomplets were displayed in sequential fashion before the participant was returned to the neighborhood.

Each session consisted of three rounds: 20 minutes for Round 1 and 10 minutes for Rounds 2 and 3. Other than these time limits, all rounds were identical in form. The extended duration of Round 1 was meant to give participants time to

familiarize themselves with both the game's keyboard controls and the game world layout. At the same time, Rounds 2 and 3 were shortened to increase the urgency of search with the aim of revealing selection biases in the participant's behavior. If a participant visited all the houses in every neighborhood before reaching the round time limit, the round would end early. We observed that 29 participants in Round 1, and one participant in Rounds 2 and 3 visited every house, ending the respective round early.

Measures and Analysis

To measure participants' overall game performance, we calculated their *effective reward* rate, R_e, that measures the total amount of reward collected per unit of linear distance traveled by the participant.

$$R_e = \frac{\sum R_{house}}{\sum \Delta d_{j-i}} = \frac{R_T}{D_T} \tag{1}$$

Here R_t is the running total of collected rewards (i.e., correctly identified Oomplets) up until the current moment in the round and D_T is the running total of distances between house visits, Δd_{j-i}, up until the current house. By using effective reward, we were able to observe performance improvements in both the exploration and selection tasks, and account for different time limits across the different rounds. In the case of the first patch visit, the coordinates of the previous patch were the center of the game world, where the player began each round. In cases where the patches were located in separate neighborhoods, the players had to travel back to the center of the arena before entering a new neighborhood. Hence, the distance between the houses in separate neighborhoods was calculated as a sum of distances from each house to the center point.

Perceptual learning was tracked using signal detection, d'. This measure is the difference of the z-scores for a participant's false alarm rate and hit rate.

$$d' = \Phi^{-1}(hit) - \Phi^{-1}(FA) \tag{2}$$

False alarm (FA) rate was computed the proportion of total selection task events where a sweet Oomplet was shown, and bitter was the incorrectly selected category. The hit rate was the proportion of these events where a bitter Oomplet was shown, and bitter was correctly selected. A $d' > 0$ indicates sensitivity in the intended direction, with increasingly stronger sensitivity as $d' \to 1$, a $d' \approx 0$ indicates no sensitivity, and a $d' < 0$ one would indicate sensitivity in the opposite of the intended signal.

To measure how well participants learned the value of each neighborhood, we introduced a *neighborhood bias (nb)* score. This was calculated as:

$$nb = \sum \frac{b_n t_n}{t_T} \tag{3}$$

where b_n is the bitter probability of neighborhood n, t_n is the time spent in neighborhood n, and t_T is the total round time. This weighs each neighborhood probability value against the time spent in the corresponding neighborhood and sums these weighted probabilities, resulting in a value between 0 and 1. *nb* was observed to be highest in rounds where a participant spent most of their exploring time in a high-value neighborhood.

To measure how well participants learned to detect the within-neighborhood spatial predictor of bitter Oomplet probability, we calculated a Spearman rank-order correlation coefficient ρ_S between the order that a participant selected a house in a neighborhood $rank_{select}$ and that house's rank order of bitter Oomplet probability $rank_{house}$:

$$hb = \rho_S(rank_{select}, rank_{house}) \tag{4}$$

We refer to this as the *house bias (hb)* score. The magnitude of this value indicated how strictly a participant adhered to the optimal exploration pattern inside of a neighborhood. *hb* is highest in cases where the highest probability house patch that is currently unexplored and unharvested is always prioritized.

All analyses of learning effects across rounds and associations between measures were performed using linear mixed-effects regression models. Participants were used as the random effects term.

Author contributions statement

Data availability

The code and assets used to create and build the BitterBuster game are available on GitHub: (https://github.com/CoAxLab/BitterBusterForagingPaperRelease).

References

1. Stephens, D. W. & Krebs, J. R. *Foraging theory*, vol. 6 (Princeton university press, 1986).

2. Kamil, A. & Roitblat, H. L. The ecology of foraging behavior: implications for animal learning and memory. (1985).

3. Schoener, T. W. A brief history of optimal foraging ecology. In *Foraging behavior*, 5–67 (Springer, 1987).

4. Barack, D. L. What is foraging? *Biol. & Philos.* **39**, 3 (2024).

5. Reynolds, A. M. & Rhodes, C. J. The lévy flight paradigm: random search patterns and mechanisms. *Ecology* **90**, 877–887 (2009).

6. Charnov, E. L. Optimal foraging, the marginal value theorem. *Theor. population biology* **9**, 129–136 (1976).

7. Niv, Y., Joel, D., Meilijson, I. & Ruppin, E. Evolution of reinforcement learning in uncertain environments: A simple explanation for complex foraging behaviors. *Adapt. Behav.* **10**, 1 (2002).

8. Kolling, N. & Akam, T. (reinforcement?) learning to forage optimally. *Curr. opinion neurobiology* **46**, 162–169 (2017).

9. Morimoto, J. Foraging decisions as multi-armed bandit problems: Applying reinforcement learning algorithms to foraging data. *J. theoretical biology* **467**, 48–56 (2019).

10. Yoon, T., Jaleel, A., Ahmed, A. A. & Shadmehr, R. Saccade vigor and the subjective economic value of visual stimuli. *J. neurophysiology* **123**, 2161–2172 (2020).

11. Dubois, T., Pasquaretta, C., Barron, A. B., Gautrais, J. & Lihoreau, M. A model of resource partitioning between foraging bees based on learning. *PLOS Comput. Biol.* **17**, e1009260 (2021).

12. Goldshtein, A. *et al.* Reinforcement learning enables resource partitioning in foraging bats. *Curr. Biol.* **30**, 4096–4102 (2020).

13. Sukumar, S., Shadmehr, R. & Ahmed, A. A. Effects of reward and effort history on decision making and movement vigor during foraging. *J. Neurophysiol.* **131**, 638–651 (2024).

14. Constantino, S. M. & Daw, N. D. Learning the opportunity cost of time in a patch-foraging task. *Cogn. Affect. & Behav. Neurosci.* **15**, 837–853 (2015).

15. Giammarino, V. *et al.* Combining imitation and deep reinforcement learning to human-level performance on a virtual foraging task. *Adapt. Behav.* **32**, 251–263 (2024).

16. Wispinski, N. J. *et al.* Adaptive patch foraging in deep reinforcement learning agents. *arXiv preprint arXiv:2210.08085* (2022).

17. Richardson-Harman, N., Phelps, T., McDermott, S. & Gunson, A. Use of tactile and visual cues in consumer judgments of apple ripeness. *J. sensory studies* **13**, 121–132 (1998).

18. Tolman, E. C. Cognitive maps in rats and men. *Psychol. review* **55**, 189 (1948).

19. Bella-Fernández, M., Suero Suñé, M. & Gil-Gómez de Liaño, B. Foraging behavior in visual search: A review of theoretical and mathematical models in humans and animals. *Psychol. research* **86**, 331–349 (2022).

20. Kasarda, J. P., Verstynen, T. & Tarr, M. J. The oomplet dataset toolkit: A flexible and extensible system for large-scale, multi-category image generation. *Sci. Reports* (2025).

21. Kacelnik, A. Central place foraging in starlings (sturnus vulgaris). i. patch residence time. *The J. Animal Ecol.* 283–299 (1984).

22. McCall, J. J. Economics of information and job search. *The Q. J. Econ.* **84**, 113–126 (1970).

23. Jacobs, E. A. & Hackenberg, T. D. Humans' choices in situations of time-based diminishing returns: effects of fixed-interval duration and progressive-interval step size. *J. experimental analysis behavior* **65**, 5–19 (1996).

24. Kolling, N., Behrens, T. E., Mars, R. B. & Rushworth, M. F. Neural mechanisms of foraging. *Science* **336**, 95–98 (2012).

25. Wolfe, J. M., Cain, M. S. & Alaoui-Soce, A. Hybrid value foraging: How the value of targets shapes human foraging behavior. *Attention, Perception, & Psychophys.* **80**, 609–621 (2018).

26. Stephens, D. W. Decision ecology: foraging and the ecology of animal decision making. *Cogn. Affect. & Behav. Neurosci.* **8**, 475–484 (2008).

27. Doeller, C. F., King, J. A. & Burgess, N. Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proc. Natl. Acad. Sci.* **105**, 5915–5920 (2008).

28. Gershman, S. J. & Daw, N. D. Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annu. review psychology* **68**, 101–128 (2017).

29. Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human orbitofrontal cortex represents a cognitive map of state space. *Neuron* **91**, 1402–1412 (2016).

30. Kumaran, D., Hassabis, D. & McClelland, J. L. What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends cognitive sciences* **20**, 512–534 (2016).

Supplementary Information

Supplementary Table 1. The effects of learning over the course of three rounds of game-play on Effective Reward

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	0.061	0.006	9.612	0.000	0.049	0.074
Round	0.020	0.003	7.349	0.000	0.014	0.025
Group Var	0.000	0.006				

Supplementary Table 2. The effects of learning over the course of three rounds of game-play on Perceptual Sensitivity (d')

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	0.914	0.221	4.138	0.000	0.481	1.346
Round	0.716	0.070	10.252	0.000	0.579	0.853
Group Var	1.559	0.516				

Supplementary Table 3. The effects of learning over the course of three rounds of game-play on Neighborhood Bias

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	0.453	0.022	20.746	0.000	0.410	0.496
Round	0.060	0.010	6.057	0.000	0.041	0.080
Group Var	0.001	0.010				

Supplementary Table 4. The effects of learning over the course of three rounds of game-play on House Bias

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	0.654	0.044	14.818	0.000	0.567	0.740
Round	0.002	0.016	0.110	0.913	-0.029	0.033
Group Var	0.046	0.073				

Supplementary Table 5. Mixed Linear Model Regression Results

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	-0.041	0.013	-3.260	0.001	-0.066	-0.016
Round	-0.001	0.002	-0.427	0.669	-0.005	0.003
DPrime	0.014	0.004	3.231	0.001	0.005	0.022
NeighborhoodBias	0.200	0.023	8.667	0.000	0.155	0.246
DPrime:NeighborhoodBias	-0.003	0.007	-0.419	0.675	-0.017	0.011
Group Var	0.000	0.004				