# Control over patch encounters changes foraging behaviour

Sam Hall-McMaster[1,2,*], Peter Dayan[3,4], Nicolas W. Schuck[1,2]

[1]Max Planck Institute for Human Development, Berlin, Germany

[2]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany and London, United Kingdom

[3]Max Planck Institute for Biological Cybernetics, Tübingen, Baden-Württemberg, Germany

[4]University of Tübingen, Tübingen, Germany

*Correspondence: hall-mcmaster@mpib-berlin.mpg.de

**Summary**

Foraging is a common decision problem in natural environments. When new exploitable sites are always available, a simple optimal strategy is to leave a current site when its return falls below a single average reward rate. Here, we examined foraging in a more structured environment, with a limited number of sites that replenished at different rates and had to be revisited. When participants could choose sites, they visited fast-replenishing sites more often, left sites at higher levels of reward, and achieved a higher net reward rate. Decisions to exploit-or-leave a site were best explained with a computational model estimating separate reward rates for each site. This suggests option-specific information can be used to construct a threshold for patch leaving in some foraging settings, rather than a single average reward rate.

## Introduction

Decision making requires anticipating how our choices will influence our fortunes in the future. Much work in neuroscience has therefore focused on how decision making can be understood as an optimisation problem based on a Markov Decision Process (MDP) formalism, in which future environmental states and rewards are under partial control of the decision maker (Sutton & Barto, 1998). This approach can account for behaviour in complex environments, where planning is needed to form an optimal policy (Huys et al., 2012; Huys et al., 2015; Kurth-Nelson et al., 2016), and can explain neural activity found in dopaminergic areas during value prediction (Schultz et al., 1997).

A common decision problem for animals in natural environments is foraging, where animals must decide between exploiting the option in front of them or leaving to search for a better alternative. Canonical Optimal Foraging Theory (OFT) considers a simplified version of this problem, in which options within an environment are encountered at fixed rates (Charnov, 1976). Under these conditions, the average reward rate in an environment is the sole variable needed to anticipate future reward outcomes for deciding to leave an option. This results in a simple optimal policy called the Marginal Value Theorem (MVT, Charnov, 1976), wherein animals should leave their current option when its reward falls below the average reward rate for the environment. Empirical studies have shown that leave decisions in mice, rats, monkeys and humans are consistent with this decision rule (Constantino & Daw, 2015, Hayden et al., 2011; Kane et al., 2017; Le Heron et al., 2020; Lottem et al. 2018). Evidence has also suggested that the unique structure of foraging decisions might be solved using neural substrates that are at least partially distinct from those involved in other reward-based decisions (Kolling et al., 2012; Rushworth et al., 2011), with the anterior cingulate cortex serving a critical role in regulating decisions about whether to disengage from an option (Fouragnan et al., 2019; Hayden et al., 2011; Wittmann et al., 2016, see Kolling et al., 2016; Mobbs et al., 2018; Rushworth et al., 2011 for reviews).

While canonical foraging decisions might not require the full algorithmic complexity of MDP solutions and could be solved using distinct brain areas from other forms of decision making (Kolling et al., 2012; Rushworth et al., 2011), more general foraging choices might not be as simple as MVT assumes. For example, options that animals encounter are not just determined by their frequency in an environment, but through decisions animals make about

which options are worth exploiting (Kamil, 1978; Merkle et al., 2014; Passingham, 1985; Sayers & Menzel, 2012; Sweis et al., 2018). In addition, resources in natural environments can replenish following exploitation (Garcia & Eubanks, 2019) and be revisited at later time points (Erwin, 1985, Lihoreau et al., 2010; Merkle et al., 2014; Seidel & Boyce, 2015). This introduces a new dilemma about when return is appropriate that is not considered under MVT because it assumes little to no revisitation (Charnov, 1976; Possingham & Houston, 1990). These characteristics raise important questions about what drives leave decisions when encounters with options are under the decision maker's control and when options promise different rewards due to their growth dynamics. Are decisions to leave an option still based on global reward information, such as the average reward rate across options in the environment, or are these choices made using local reward information about specific alternatives?

We recently proposed that foraging-inspired tasks used in neuroscience could be made more ecological by allowing animals to choose what option is encountered next, after deciding to leave a current option (Hall-McMaster & Luyckx, 2019). In the present experiment, we therefore studied foraging decisions in an environment where foraging choices could be directed towards specific options and option resources replenished over time. As compared to situations where options are encountered at random, we predicted that decision makers would use their knowledge about replenishment to visit faster replenishing options more often, increasing the average reward per action. We also predicted that decisions to leave an option would be made at higher reward levels. Based on MVT, this change in leaving threshold would be expected due to a higher global reward rate. However, an alternative mechanism would be that decision makers leave based on option-specific reward estimates when their next option can be selected.

In the present study, we tested the predictions above and sought to address how global and local reward information are used to decide when to leave an option. Human participants completed a patch-leaving task similar to studies from Optimal Foraging Theory (Stephens & Krebs, 1986), in which an option's reward decreased over time through exploitation. In contrast to previous studies (Constantino & Daw, 2015, Hayden et al., 2011; Kane et al., 2017; Le Heron et al., 2020), participants switched between the same three options throughout a block, which replenished their rewards at distinct rates when not being exploited. Participants therefore needed to decide when to leave their current option in order to revisit and exploit one of the alternatives. Between blocks, we manipulated whether participants had a free choice over which option to

exploit after deciding to leave the current option or whether the choice was forced, being randomly determined for them.

When participants were able to control what options were encountered, we found that fast replenishing options were visited more often than slow replenishing options. This increased the average reward rate and elevated leaving thresholds, with higher reward outcomes prior to leave decisions. Due to higher reward both when arriving at options and when leaving them, no difference was observed in the number of exploit decisions before leaving, between free and forced choice conditions. Participants' exploit-or-leave decisions were best explained with a computational model that tracked separate reward rates for each option and decided to leave the current option when its reward fell below the reward rate of the best alternative.

## Results

To investigate the effect of decision control on foraging behaviour, 60 participants completed an online foraging task. Participants controlled a pirate ship, sailing to three equidistant islands to dig for buried treasure. Participants first selected an island to sail to and then made a series of exploit-or-leave decisions. When deciding to exploit, participants dug for treasure and received between 0 and 100 gold coins based on the island's current reward stocks. The number of coins that were unearthed depleted exponentially across successive exploit actions on the island. While the current island was being exploited, new treasure was buried at each alternative island at an island-specific rate (slow, medium, fast). When deciding to leave the current island, participants sailed to one of the two alternative islands. This structure repeated until participants had performed a total of 200 actions and the block ended. Choosing islands to visit, exploiting islands and leaving all counted as individual actions. Sailing did not consume the actions available in a block. The central manipulation in this task was whether participants had a free or forced choice over the island they would sail to next, after deciding to leave the current island. In free choice blocks, the seas were calm and participants could select either alternative island, following a leave decision. In forced choice blocks, the seas were stormy and islands could not always be reached. Participants were therefore forced to sail to just one of the alternative islands, determined at random, following each leave decision. The island that would become available was only revealed after the leave decision was made. Participants completed two free choice blocks and two forced choice blocks in a random order, being informed about the upcoming condition prior to each block. Islands were located at one of three vertex positions that formed an equilateral triangle on screen, but slow, medium and fast replenishing islands were assigned at random to different positions for each block. Participants were instructed that blocks would contain a fast, slow and medium replenishing island but did not have direct experience with the islands before starting the experiment. In the results that follow, we refer to each island as a patch, consistent with terms used in Optimal Foraging Theory (Stephens & Krebs, 1986; Charnov, 1976).

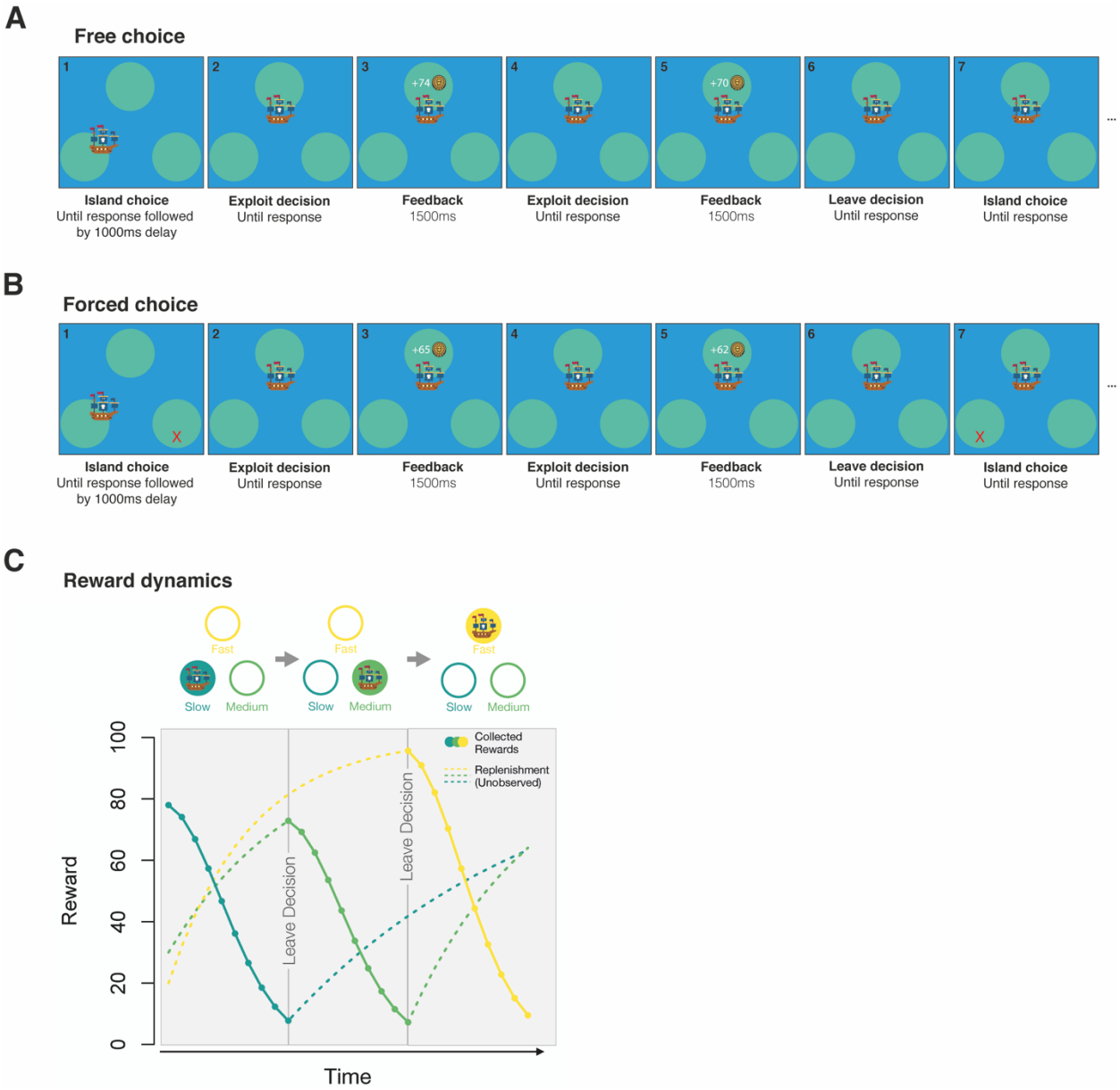


*Figure 1.* Example trial sequences for free and forced choice conditions. A: Free choice trials. Participants first chose an island to sail to (screen 1). Once their ship had arrived, participants could dig on the island for buried treasure (an exploit action shown on screen 2). Following a dig, participants received feedback about the number of coins added to their treasure chest (screen 3). Participants could continue digging (screens 4-5), until deciding to leave the island (screen 6) and choosing another island to sail to (screen 7). B: Forced choice trials. Under forced choice conditions, the seas were stormy and islands were not always accessible. Participants were forced to choose just one of the two alternative islands, after a leave decision. The accessible island was randomly determined after each leave decision and the other, inaccessible island was marked with a red X. All other aspects were as in the free choice condition shown in panel A. C: Reward dynamics. When an island was being scoured for treasure, the number of coins added to the treasure chest decreased with each successive dig. The depletion rate during exploitation was the same for all islands. While an island was being exploited, new coins were

buried on the alternative island at island-specific rates (slow, medium or fast, shown in sea green, green and yellow lines respectively). Participants did not see new coins being buried and therefore needed to revisit islands to learn about their replenishment speeds. No reward was given for choosing an island to sail to or deciding to leave an island (not indicated in panel).

**Participants directed patch visits in proportion to expected reward**

We first examined how often participants decided to visit each patch under free and forced choice conditions. In line with our predictions, we found a significantly higher proportion of visits to the fast replenishing patch in free compared with forced choice blocks (mean free choice=0.366, SD=0.027; mean forced choice=0.328, SD=0.037; $t(59)=-5.946$, $p=1.588e-7$), and a significantly lower proportion of visits to the slow replenishing patch in free choice blocks (mean free choice=0.289, SD=0.035; mean forced choice=0.341, SD=0.035; $t(59)=7.179$, $p=1.346e-9$, see Figure 2A). The proportion of visits to the medium replenishing patch was also significantly higher in free choice blocks (mean free choice=0.345, SD=0.025; mean forced=0.331, SD=0.037; $t(59)=-2.467$, $p=0.017$). While proportions sum to one within each condition in the tests above, logit transforming values to approximate normal distributions showed the same results (fast patch free vs forced $t(59)=-2.289$, $p=0.027$; slow patch free vs forced $t(59)=5.937$, $p=1.646e-7$).

Exploratory analyses revealed that participants selected the alternative patch with the higher expected reward significantly more often under free choice than forced choice conditions. This effect was observed when leaving the slow patch (mean free=0.711, SD=0.195; mean forced=0.511, SD=0.121; $t(59)=-6.783$, corrected $p=1.893e-8$), the medium patch (mean free=0.615, SD=0.178; mean forced=0.480, SD=0.113; $t(59)=-5.090$, corrected $p=1.176e-5$) and the fast patch (mean free=0.641, SD=0.166; mean forced=0.493, SD=0.135; $t(59)=-5.520$, corrected $p=2.392e-6$, see Figure 2B). When considering choices based on patch replenish rates, rather than expected rewards, participants selected the alternative patch with the higher replenishment rate significantly more when leaving medium replenishing patch under free choice conditions (mean free choice=0.615, SD=0.171; mean forced choice=0.455, SD=0.134; $t(59)=-5.022$, corrected $p=1.509e-5$). Significant differences were not detected in how often the higher replenishment rate patch was selected when leaving the fast replenishing patch (mean free choice=0.570, SD=0.189; mean forced choice=0.510, SD=0.112; $t(59)=-2.105$, corrected

$p$=0.119) or the slow replenishing patch (mean free choice=0.539, SD=0.196; mean forced choice=0.510, SD=0.125; $t(59)$=-0.928, corrected $p$>0.99).

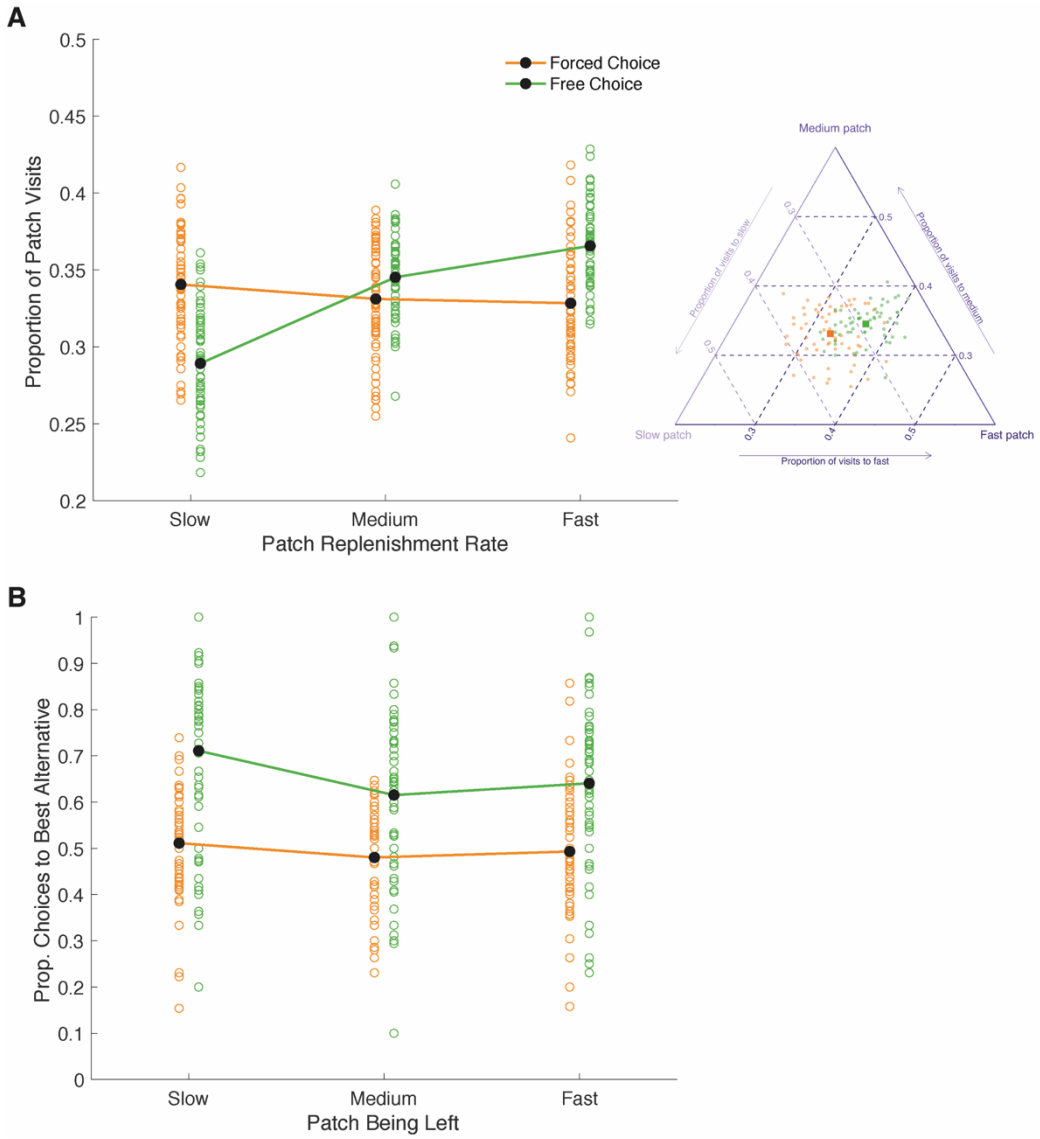


*Figure 2.* Patch choices. A: Proportion of visits to slow, medium and fast replenishing patches, under forced and free choice conditions. Orange colours indicate data from forced choice blocks and green colours indicate data from free choice blocks. In the left subpanel, unfilled circles show proportions for individual participants and lines connect average proportions across the sample. The right subpanel shows the proportion of visits to each patch as a single point on a simplex for each participant. Squares indicate mean proportions for each choice condition. For both plots in panel A, proportions across the three patches sum to one within each choice condition. B: Proportion of choices to the patch with the highest expected reward, when leaving the slow, medium and fast replenishing patches. Colours are the same as those used in panel A.

**Reward rates were higher when patch encounters could be directed**

The ability to direct foraging towards patches with faster replenishing dynamics should result in higher rewards overall and, therefore, increase experienced reward rates. In line with this prediction, we observed that the reward per exploit action was significantly higher in free compared to forced choice conditions (mean=69.103 vs. 64.808 respectively; SD free choice=9.326, SD forced choice=8.684; $t(59)=-6.527$, $p=1.699e-8$, see Figure 3A). This pattern did not change when using the average reward rates across all actions (including patch choices, exploit actions and leave actions), which was significantly higher in the free (mean=48.239, SD=2.727) compared with forced choice blocks (mean=45.929, SD=2.761; $t(59)=-7.970$, $p=6.155e-11$, see Figure 3B). Exploratory analyses found that the reward per exploit action was higher for all patch types under free choice conditions, including the slow replenishing patch (mean free choice=65.003, SD free choice=10.158; mean forced choice=59.402, SD forced choice=9.616, $t(59)=-6.635$, corrected $p=3.357e-8$), the medium replenishing patch (mean free choice=68.958, SD free choice=9.171; mean forced choice=65.093, SD forced choice=8.636; $t(59)=-5.379$, corrected $p=4.048e-6$) and the fast replenishing patch (mean free choice=70.902, SD free choice=9.008; mean forced choice=67.861, SD forced choice=8.149; $t(59)=-5.239$, corrected $p=6.810e-6$).

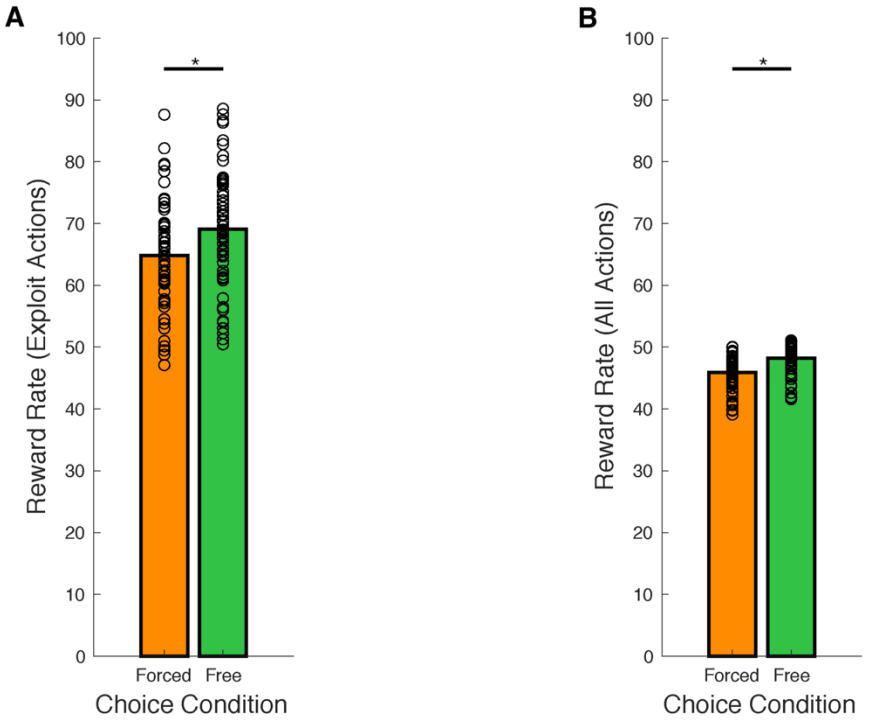


*Figure 3.* Reward per action (reward rate) under free and forced choice conditions. A: Reward rates based on exploit actions. B: Reward rates based on all actions, which includes selecting patches to visit, exploit actions and leave actions. Coloured bars show mean reward rates for each condition. Circles overlaid on each bar show individual participant values. *p<0.05.

**Reward threshold for leave decisions was higher when patch encounters could be directed**

Optimal Foraging Theory (Charnov, 1976) predicts that as experienced reward rates increase, patches will be abandoned at higher reward values. The previous result, showing higher reward rates in free compared to forced choice blocks, therefore suggests there might be differences in the average reward value before leaving. To test this idea, we examined the last reward outcome before leaving each patch under free and forced conditions. We observed a significant main effect of choice condition ($F(1,59)=18.723$, p=5.923e-5), reflecting higher reward outcomes prior to leave decisions in free (mean=48.753, SD=18.070) compared with forced choice blocks (mean=44.418, SD=16.240, see Figure 4). There was also a significant main effect of patch type ($F(1,59)=15.376$, $p$=1.164e-6), reflecting lower reward outcomes before leaving the fast replenishing patch (mean=45.503, SD=16.686), compared to the medium replenishing patch (mean=46.728, SD=17.113; $t(59)=-3.481$, corrected $p=0.003$) and compared to the slow replenishing patch (mean=47.526, SD=16.467, $t(59)=-5.009$, corrected $p$=1.583e-5).

A significant difference was not detected between the reward before leaving the medium and slow replenishing patches ($t(59)=-2.322$, corrected $p=0.071$). No significant interaction was detected between choice condition and patch type on the reward outcome prior to leaving ($F(1,59)=0.021$, $p=0.980$).

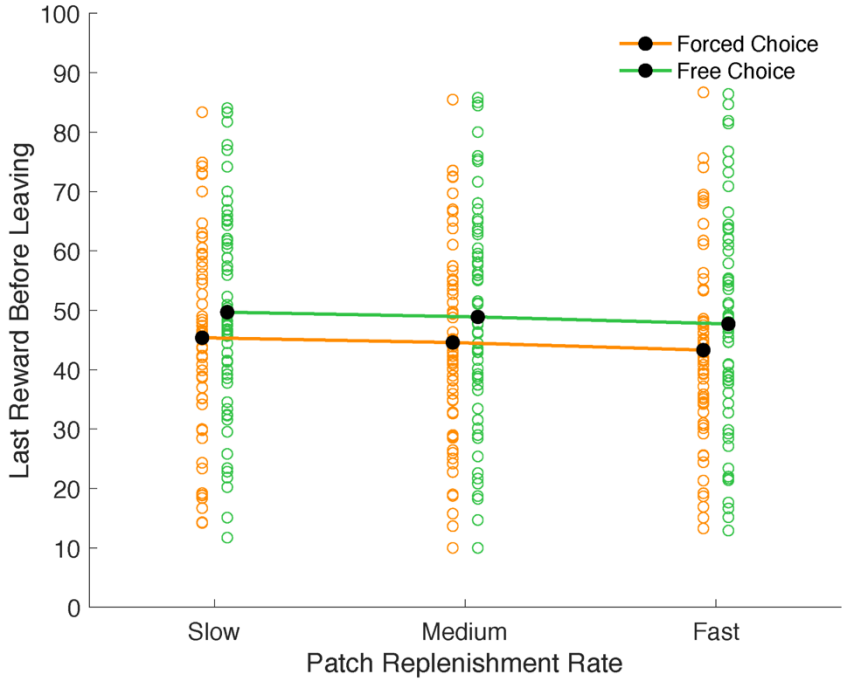


*Figure 4.* Last reward prior to leaving the slow, medium and fast replenishing patches. Orange indicates data from forced choice blocks and green indicates data from free choice blocks. Black dots indicate averages and unfilled circles show values for individual participants.

**Directed patch choices result in higher rewards on arrival and a comparable number of exploit actions before leaving**

If participants selected fast replenishing patches more often, higher reward rates might not just reflect patches being abandoned at higher reward values, but the average patch reward being higher at arrival. When investigating this possibility, we found a significant main effect of choice condition on reward at arrival to a patch ($F(1,59)=316.847$, $p=2.127e-25$), indicating that the reward upon arrival at a patch differed in free and forced choice blocks (see Figure 5A). Moreover, we observed a significant main effect of patch type on reward at arrival ($F(1,59)=675.677$, $p=2.402e-65$) and a significant interaction between choice condition and patch type ($F(1,59)=33.1867$, $p=3.671e-12$). Exploratory analyses revealed that the interaction

resulted from a smaller condition difference in reward when arriving at the fast replenishing patch (mean free=91.874 vs mean forced=90.600; SD free=1.830, SD forced=1.863; $t(59)=-4.653$, corrected $p=5.695e-5$) and increasing condition differences in reward when arriving at the medium replenishing patch (mean free=87.271 vs mean forced=84.575; SD free=2.287, SD forced=2.494; $t(59)=-7.040$, corrected $p=6.941e-9$) and the slow replenishing patch (mean free=79.421 vs mean forced=72.869; SD free=3.613, SD forced=4.550; $t(59)=-11.166$, corrected $p=1.054e-15$).

A related set of analyses, examining arrival rewards as a function of the patch being left, showed a significant main effect of choice condition on the arrival reward ($F(1,59)=153.055$, $p=4.928e-18$), reflecting higher arrival rewards in the free choice condition (mean free=80.470, SD=2.340; mean forced=76.036, SD=3.514, see Figure 5B). There was also a significant main effect of the patch being left on the reward gained when arriving at a new patch ($F(1,59)=60.296$, $p=9.105e-19$). Exploratory analyses showed that this main effect was due to higher arrival rewards after leaving the slow compared to the medium patch (mean slow=81.966, SD=2.975; mean medium=77.643, SD=4.005; $t(59)=6.605$, corrected $p=3.770e-8$), the slow compared to the fast patch (mean fast=75.150, SD=4.443; $t(59)=10.975$, corrected $p=2.106e-15$) and the medium patch compared to the fast patch ($t(59)=4.102$, corrected $p=3.822e-4$). No significant interaction was detected between the patch being left and choice condition ($F(1,59)=1.185$, $p=0.310$).

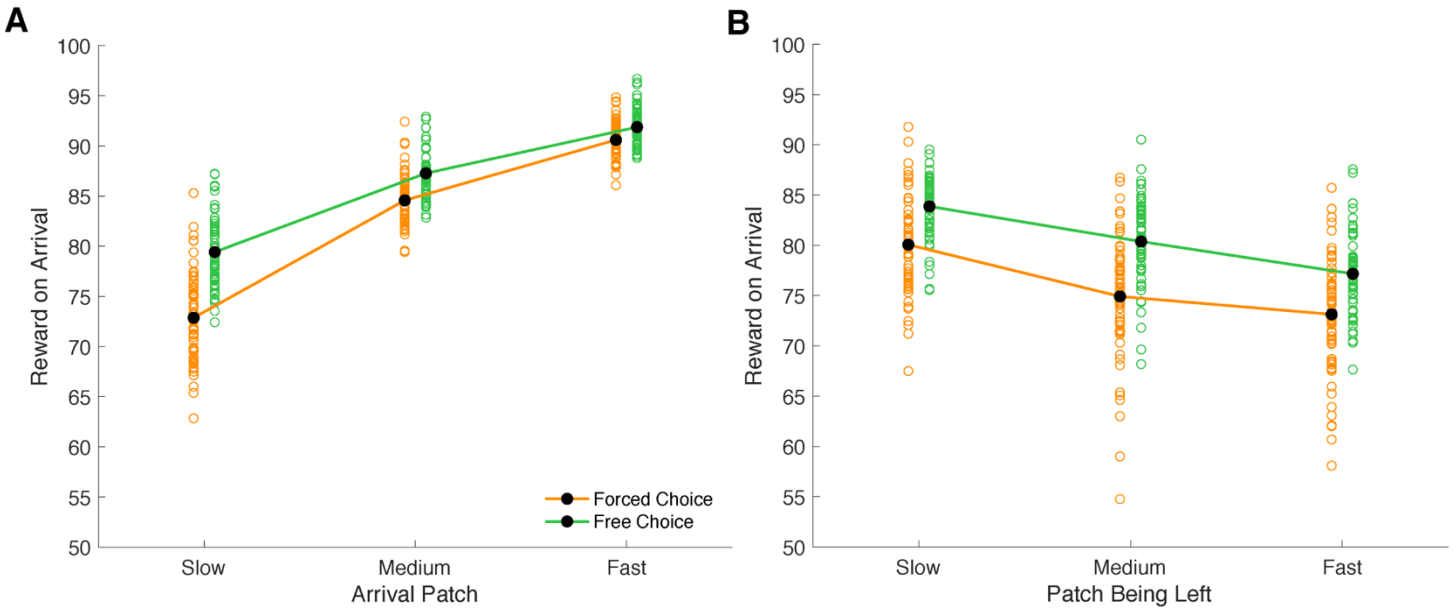


*Figure 5.* Reward on arrival. A: Average reward outcome for the first exploit action when arriving at slow, medium and fast replenishing patches. Orange indicates data from forced choice blocks and green indicates data from free choice blocks. Black dots indicate sample means and unfilled circles show individual participant values. B: Average reward on arrival at a new patch, after leaving the slow, medium and fast replenishing patches. Colours are the same as those used in panel A.

The combined effects of higher rewards at arrival and before leaving in the free compared to the forced choice condition resulted in a similar number of exploit actions before patch leaving across conditions (mean free=5.276, SD=1.786; mean forced=5.424, SD=1.659; $F(1,59)=3.337$, $p=0.073$, see Figure 6). There was a significant main effect of patch type ($F(1,59)=336.6414$, $p=1.736e-49$) and a significant interaction between patch type and choice condition on the number of exploit actions before leaving ($F(1,59)=9.312$, $p=1.758e-4$). Exploratory analyses revealed that the interaction resulted from significantly fewer exploit actions prior to leaving the fast replenishing patch in the free choice condition (mean free choice=5.672, SD free choice=1.753; mean forced choice=6.015, SD forced choice=1.554; $t(59)=3.400$, corrected $p=0.004$), but no condition differences in leaving times for the slow replenishing patch (mean free choice=4.773, SD free choice=1.798; mean forced choice=4.759, SD forced choice=1.787; $t(59)=-0.103$, corrected $p>0.99$) or the medium replenishing patch

(mean free choice=5.309, SD free choice=1.850; mean forced choice=5.581, SD forced choice=1.708; $t(59)=2.207$, corrected $p=0.094$).

Together with our analyses of arrival reward rates, these results suggest that when the reward at arrival was more closely matched across choice conditions (i.e. in the fast replenishing patch), patches were abandoned after fewer actions under free choice conditions. When the reward at arrival was not well matched across choice conditions (i.e. in the slow and medium replenishing patches), the higher rewards at arrival and prior to leaving seen in free choice blocks cancelled each other out, resulting in a similar number of actions before leaving to the forced choice condition.

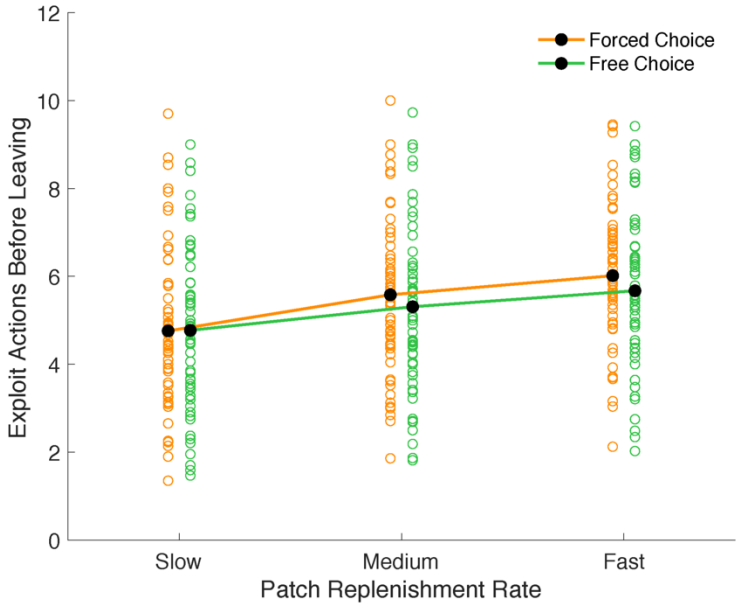


*Figure 6.* The number of exploit actions made before deciding to leave the slow, medium and fast patches, under free and forced choice conditions. Orange indicates data from forced choice blocks and green indicates data from free choice blocks. Black dots indicate averages and unfilled circles show individual participant values.

**Local reward information best accounted for exploit-leave decisions**

The results above indicate that being able to control patch encounters altered patch leaving decisions. Fast and moderate replenishing patches were chosen more frequently, increasing reward rates in the free choice condition. Crucially, this allowed participants to leave patches at higher reward values, suggesting a change in the core computational mechanisms that

underlie leave decisions. In particular, participants might have used information about specific options as a threshold for leaving, instead of the average reward rate across all options proposed under MVT.

To test this idea, we compared two computational models that based leave decisions on either on global reward expectations (the average reward rate across all patches) or local reward expectations (reward rates associated with patches not currently being exploited). Both models computed the probability of exploiting the current patch at time *t* using a logistic decision rule based on previous work (Constantino & Daw, 2015).

$$P(exploit)_t = \frac{1}{1 + exp(-[c + \beta(r_{t-1} - T_t)])}$$

The core aspect of the expression above is the term $r_{t-1} - T_t$ which compares $r_{t-1}$, the last reward received from the current patch at time *t-1,* to the current leaving threshold of the model $T_t$. The parameter $c$ is an intercept that estimates how much people tend to over or under exploit the current patch relative to the reward rate. $\beta$ is a parameter that controls the slope of the logistic function, reflecting participants' sensitivity to the difference between the leaving threshold and the most recent reward outcome. Both models calculated their leaving threshold *T* on the basis of their current reward rate estimate, updating their reward rate estimate following each exploit decision using a simple delta learning rule with a learning rate $\alpha$. Importantly, the local model estimated reward rates separately for each patch *p*:

$$\hat{r}_{p,t} = \hat{r}_{p,t-1} + \alpha(r_t - \hat{r}_{p,t-1}),$$

updating the reward rate estimate for the currently visited patch *p,* and leaving reward rate estimates of the other patches, *q*, unchanged:

$$\hat{r}_{q,t} = \hat{r}_{q,t-1} \; \forall q \neq p$$

At decision time, the leaving threshold was constructed as the maximum of the two estimated reward rates of two alternative (i.e. unattended) patches:

$$T_t = \max_q(\hat{r}_{q,t})$$

The global model, in contrast, kept track of a single global reward rate across patches:

$$\hat{r}_{G,t} = \hat{r}_{G,t-1} + \alpha(r_t - \hat{r}_{G,t-1})$$

and used this directly as the leaving threshold:

$$T_t = \hat{r}_{G,t}$$

Reward rate estimates were initialised at 50 at the start of each block in both models and were updated after each exploit action. When deciding to leave a patch, the relevant reward rate estimate (i.e. the global estimate in the global model and the current patch estimate in the local model), was down-weighted using the update expressions above. In this case, $r_t$ was set to 0 and a separate learning rate $\alpha_l$ was used. Both models had 4 free parameters, $\alpha$, $\alpha_l$, $\beta$ and $c$.

When comparing global and local models, we found that the local model better explained exploit-or-leave choices in the free choice condition ($BF_{10}$=6.016; mean BIC difference=6.102; participants best fit in free choice condition with local reward rate model=44 and global model=16, see Figure 7B). Despite large behavioural differences between free and forced choice blocks, the local model also performed better than the global model in the forced choice condition ($BF_{10}$=18.353, mean BIC difference=9.257, participants best fit in forced choice condition with local model=37 and global model=23, see Figure 7A). While the local reward rate model best accounted for more participants in the free choice condition, we did not find evidence that it provided a better account of choice behaviour between conditions ($BF_{10}$=0.412, mean BIC difference free vs forced=9.354). The maximum estimated reward rate among the alternative patches in the local model showed a substantial correlation with the mean over the alternative reward rate estimates in both free choice (mean rho=0.910) and forced choice blocks (mean

rho=0.893). While both such thresholds are constructed based on local reward rate estimates, the precise information used as the leaving threshold is difficult to arbitrate in the current setup.

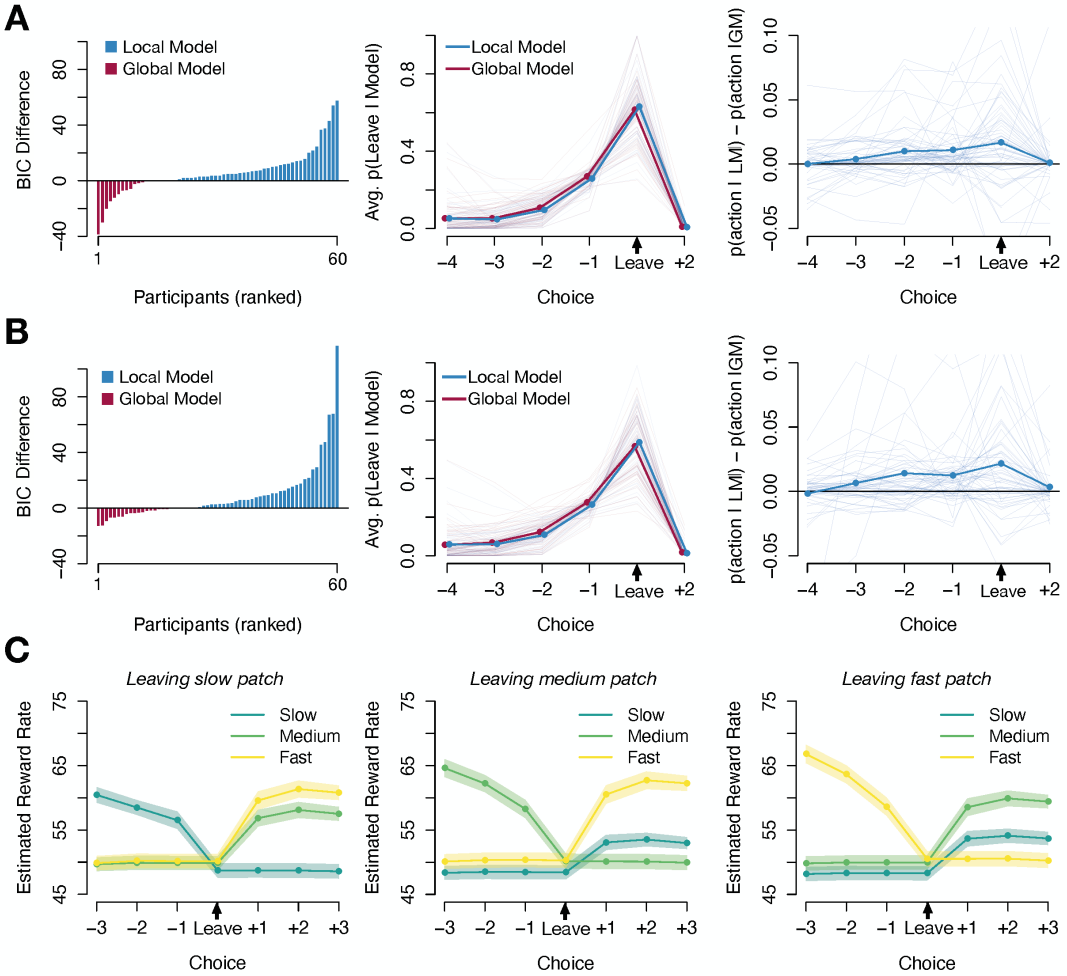


*Figure 7.* Modelling results for exploit-leave decisions. A: Forced choice results. The left subpanel shows differences in Bayesian Information Criterion between the global reward model (red), using a single reward rate as its leaving threshold, and the local reward model (blue), using patch specific reward rates as leaving thresholds. Higher positive BIC differences indicate the local model better accounted for exploit-leave decisions. The middle subpanel shows average probabilities of making a leave response at different time points, relative to leave decisions, for global and local models. Thin lines show individual participant data. The right subpanel shows differences in the probability for the choice participants made under global and local models. Positive values indicate better local model performance. Thin lines show individual participant data. B: Free choice results, using the same panel and color structure as A. C: Patch specific reward rate estimates under the local reward rate model when leaving the slow patch (left), medium patch (middle) and fast patch (right). Shading around the principle lines indicates standard error of the mean.

## Discussion

Foraging has often been studied as a case of decision making in which animals have little control over patches encountered in the future (Barack et al., 2017; Constantino & Daw, 2015; Hayden et al., 2011; Kane et al., 2017; Le Heron et al., 2020). This assumption might not always be true (Merkle et al., 2017, 2014; Passingham, 1985; Riotte-Lambert et al., 2015; Sayers & Menzel, 2012), complicating optimal foraging decisions. The present study aimed to investigate patch leaving decisions in an environment in which the same set of three patches could be revisited, and patches replenished their rewards at different rates. Within such an environment, we manipulated whether participants had control over patch encounters, studying how this influenced decisions about when to leave a patch and the information used to make leave decisions.

Consistent with our predictions (Hall-McMaster & Luyckx, 2019), we found that when participants were able to select which patch to exploit after a leave decision, they visited fast replenishing patches more often than slower replenishing patches. This resulted in higher reward rates, higher rewards when arriving at patches and higher reward outcomes prior to leave decisions, compared to environments in which patch encounters were randomly determined. Increased rewards at arrival and prior to leaving patches in the free choice environment resulted in a comparable number of actions before leaving patches to the forced choice environment. To account for these effects, we developed a simple computational model for exploit-or-leave decisions, which estimated patch specific reward rates and increased the probability of a leave decision when reward from the current patch fell below the highest reward rate among the alternative patches. We found that choices in free choice blocks were better explained with this local reward rate model than an MVT-based model that estimated a single, global reward rate. Despite large behavioural differences between choices conditions, decisions in forced choice blocks were also better captured using the local model.

The present results build on previous studies based on Optimal Foraging Theory (Charnov, 1976; Stephens & Krebs, 1986), which involved random or pseudo-random encounters with new options when deciding to leave a current option (Barack et al., 2017; Constantino & Daw, 2015; Hayden et al., 2011; Kane et al., 2017; Le Heron et al., 2020). Here we show that allowing decision makers to select which options are pursued results in dramatic changes to foraging behaviour. More frequent visits to high value options and less frequent visits

to low value options increase the reward accrued per action, reflecting behaviour that is more efficient than encountering options at random and could improve biological fitness in natural settings (Merkle et al., 2017, 2014; Riotte-Lambert et al., 2015; Sayers & Menzel, 2012). This increased efficiency alters the reward threshold for leaving current options, seen in the present experiment as higher reward outcomes prior to leave decisions when patches could be freely selected. Results showing that exploit-or-leave choices were better explained with a local reward rate model suggest that humans use reward information about specific alternatives to construct leaving thresholds, in at least some foraging environments. The strong behavioural differences between choice conditions indicate participants adapted their behaviour to different foraging constraints, but not in a manner that was explained with shifts between global and local patch leaving models. This does not necessarily indicate that participants used the same information for leaving in free and forced choice blocks; there could still be core computational differences between choice conditions that were not captured with the models tested here, such as greater prospective tracking of replenishing patch rewards under free choice conditions. Based on the high correlation between the maximum reward rate among the alternatives and the mean across the alternatives in the local model, the precise information used as a leaving threshold is still unclear. Critically, both potential thresholds require separate reward rate estimates for each patch, which either get filtered or transformed. The present results therefore lend credence to the proposal that local reward information is used in some foraging environments to construct the leaving threshold, rather than the average reward rate for the environment proposed under MVT (Charnov, 1976).

Collectively, the results from this experiment suggest that additional aspects of real-world decision making, such as control over the options we pursue and revisiting options, influence decisions about when to leave options and what reward information is used to construct a threshold for leaving.

## Limitations of the Study

The present experiment has three main limitations. First, the experiment did not provide a normative account for optimal patch leaving in environments where three patches can be revisited and agents have control over patch encounters (see Kilpatrick et al., 2020; Possingham & Houston, 1990 for two-patch revisiting solutions). While foraging patterns showed striking differences between choice conditions with respect to the patches visited, rewards on arrival, rewards prior to leaving and net reward rates, we have not assessed whether these would be the same differences shown under an optimal solution. Second, the experiment did not provide a computational account for how patches were selected. Participants tended to select patches with higher expected rewards when leaving all three patches and higher reward rate patches when leaving the medium replenishing patch. The local reward rate model assumed that participants used the alternative patch with the highest reward rate estimate as a threshold for leaving, but future efforts should model both exploit-leave decisions and patch selection decisions, to understand how patch selection processes interact with the information used to make leave decisions. Moreover, participants did not know the replenishment speed for each island at the start of a block or the precise rates of replenishment. Behaviour at the beginning of a block might therefore reflect directed exploration, which is not accounted for in the models we present. Third, the experiment could be limited in its trial-based structure. While some patch-leaving designs are trial-based (Barack et al., 2017; Constantino & Daw, 2015; Kane et al., 2017; Hayden et al., 2011), studies tend to compute reward rates (Kane et al., 2017; Lottem et al., 2018), the probability of leaving (Constantino & Daw, 2015) or leaving times (Barack et al., 2017; Hayden et al., 2011; Le Heron et al., 2020; Lottem et al., 2018) using some continuous time information, such as the number of seconds in a patch. In the present experiment, we used trial-wise data to model discrete choice behaviour based on a previous computational model (Constantino & Daw, 2015). Unlike Constantino and Daw's (2015) task, however, travel times were fixed in the present experiment and did not reduce the number of actions available in a block. Decisions about which patch to visit did count towards the number of actions in a block and travel time here refers to a 1s delay between patch selection and arriving at a patch. We therefore removed continuous temporal variables from Constantino and Daw's original (2015) model, which were used to capture the exploit time and travel time in seconds. While appropriate

for our task design, this could have resulted in less sensitive reward rates and leaving time estimates, as compared with a continuous time-based setup.

## Method

### Participants

70 participants took part in this experiment as an online study, which ran using Prolific (https://www.prolific.co/). Participants were eligible to take part if they were 18-40 years of age, fluent in English and were not receiving treatment or taking medication for mental illness. Eligible participants were screened with attention checks before participation was allowed. Ten participants, from the 70 who passed the attention checks and completed the experiment, were excluded due to their task performance (number of points earned) falling more than three median absolute deviations below the sample median. The remaining 60 participants were between 18 and 40 (mean age=28, 33 female, 27 male). Participants received £5 per hour and could earn up to £1 extra for depending on their task performance. The study was approved by the Max Planck Institute for Human Development Ethics Committee (N-2020-06) and all participants indicated their informed consent before taking part.

### Materials

Stimuli were presented using Psychopy-3 (https://www.psychopy.org/). Stimuli included three circles, which were used as different choice options, a cartoon pirate ship, and a cartoon medallion, used for reward feedback (see Figure 1). Cartoon stimuli were created by icon developers Smalllikeart and Nikita Golubev, and accessed at https://www.flaticon.com/. Stimulus sizes were scaled based on participants' screen sizes and responses were recorded using participants' keyboards. Participants were asked to only take part if their screen was at least 13 inches and they were using Chrome as their browser. The task was hosted on Pavlovia (https://pavlovia.org/) which stored participant data during online testing.

### Code Accessibility

All materials needed to run the task online, experimental data and analysis code can be accessed at: [link will be added on publication].

### Procedure

Participants performed a patch-leaving task, based on principles from Optimal Foraging Theory (Stephens & Krebs, 1986). The classic patch-leaving setup involves participants

exploiting an option as the reward it returns decreases over time. Participants' central choice is when the option is no longer worth exploiting; when it should be abandoned to search for something better. Once participants decide to leave the option, a new option is presented for them to exploit, the value of which is independent of their previous actions. We made three critical adaptations to this standard setup. First, we manipulated whether participants could choose their next option, after a leave decision. Second, participants revisited the same three options within a block, which meant that their previous decisions would influence an option's reward value when returning to it. Third, each option replenished its reward over time, but at a different rate.

Participants played the role of a pirate, sailing to different islands in search of buried treasure (Figure 1). The common elements across all conditions of the task were as follows. Participants first selected an island to sail to using the keys F (left island), J (upper island) and K (right island). Following a travel delay of 1000ms to reach the island, participants made a series of exploit-leave decisions. When deciding to exploit the island for treasure (space bar press), participants received between 0 and 100 gold coins, which were indicated on screen for 1500ms. The reward feedback then disappeared and participants made another exploit-leave decision. This loop continued until the participant decided to leave the island (S key press), following which they were able to select a new island to sail, and thereafter enter a new exploit-leave loop. Participants needed to perform at least one exploit action before being able to leave an island. The event sequence described above continued until the participant had performed a total of 200 actions, at which point the block ended. Choices about which island to sail to, exploit decisions and leave decisions all counted as individual actions. The actions remaining in a block were shown in the top right corner of the screen. While sailing between islands took time, reducing the reward per unit time, it did not affect the reward per action because sailing did not consume the actions available in a block. We therefore used the reward per action as our main measure of reward rate. The critical manipulation between blocks was whether participants had a free or forced choice over which island they sailed to, after each leave decision. In free choice blocks, the seas were calm and participants could choose to visit either of the two alternative islands, after a leave decision. In forced choice blocks, the seas were stormy and islands were not always accessible. Participants were therefore forced to select one of the alternative islands, which was

randomly determined at each selection phase. Inaccessible islands were marked with red X symbols.

The reward dynamics for each island functioned the same way across task conditions. When visiting an island/patch, *p*, the number of coins gained for the first exploit action was equal to the full number of coins buried. When a subsequent exploit action was made, the reward gained at time *t,* (denoted as $r_{p,t}$ ) decreased following the equation:

$$r_{p,t} = d^{n-1} r_{p,t-1}$$

The decay constant, *d*, was set to 0.95, *n* refers the number of exploit actions since arriving at the island, and $r_{p,t-1}$ refers to the reward gained for the previous exploit action. Note that the decay constant declines exponentially in this expression, leading to accelerating decay characterized by a "squared exponent" relationship between the reward when entering the patch at time *t* and the reward after making *k* exploit actions prior to the current exploit action (i.e. *k* = *n*-1):

$$r_{p,t+k+1} = r_{p,t} \prod_{i}^{k} d^{i} = d^{0.5k(k+1)} r_{p,t}$$

While one island was being exploited, new coins were buried on the other two islands. The coins buried on each alternative island *q* increased, following each decision to exploit the current island, according to the equation:

$$r_{q,t} = r_{q,t-1} + (100 - r_{q,t-1}) k_q \quad \forall q \neq p$$

The variable $k_q$ is the replenish rate for alternative patch *q*. The replenish rate was different for each island, with one island replenishing at a slow rate (0.05), one at a moderate rate (0.10) and one at a fast rate (0.15). This equation was also used to update the coins buried on all islands, after island selection responses and leave responses. Each island could replenish to a maximum of 100 coins. In each block, the slow, medium and fast replenishing island was randomly assigned to the left, top or right position on screen.

Participants completed four blocks, two free choice blocks and two forced choice blocks. The block order was random. Participants were informed that each block would contain a fast, slow and medium replenishing island but did not have direct experience with the islands before starting the experiment. Participants were instructed about whether an upcoming block would be free or forced choice and were reminded to gain as much reward as possible. The block began with a choice between all islands (free choice) or a randomly selected island (forced choice). Islands were located at one of three vertex positions on screen that formed an equilateral triangle. However, slow, medium and fast replenishing islands were randomly assigned to the vertex positions in each block. Participants therefore needed to revisit different islands (left, top, right) to learn about their replenishment speeds when starting a new block. The coins buried at each island were initialised to random numbers between 69 and 79 at the beginning of the block. This range was determined based on simulating 1000 agents that completed one forced choice block, using the average experienced reward rate to make leave decisions. In these simulations, the initial reward for each option was set to 100. Average experienced reward rates were calculated prior to each stay-leave decision, by dividing the total reward gained by the total number of actions taken. The reward gained for the first exploit action when arriving at an island was recorded during the second half of the block, allowing time for the islands to reach steady reward dynamics. We then calculated 95% confidence intervals for the recorded values and averaged confidence intervals across agents, resulting in a lower bound of 69.3521 and an upper bound of 79.1659 points.

## Data Analysis

### *Visitation*

To test the prediction that participants would revisit high value options more and low value options less under free choice conditions, we extracted the number of times each patch (slow, medium, fast replenishment rate) was selected in forced and free choice blocks. The number for each patch in each condition was then normalised by the total number of selection actions performed in its respective condition. For example, the number of visits to the fast patch in the free choice blocks was divided by the total number of visits made to any patch in the free choice blocks. This normalisation step means that the proportion of visits to each patch in the forced choice condition sum to one and the proportion of visits to each patch in the free choice

condition sum to one. Normalised visits under free and forced choice conditions were compared for each patch using separate paired two-tailed t-tests. To address potential issues with non-normally distributed (bounded) proportions, we logit transformed values. To do so, the proportion of visits to the slow and fast patches were divided by the proportion of visits to the medium patch, separately in each choice condition. The resulting values were log transformed. The transformed values were then compared between choice conditions using separate paired two-tailed t-tests.

In exploratory analyses, we examined how often the alternative patch with the highest expected reward (number of coins buried) was chosen, when leaving each patch. For each patch being left (slow, medium, fast replenishment rate), we extracted the number of coins for each alternative patch at the time of choosing the next patch to sail to. We then computed the proportion of times participants chose the alternative patch with the highest number of coins. Proportions for each patch being left were compared between choices using separate two-tailed t-tests. Alpha thresholds were adjusted for three exploratory tests using the Bonferroni correction. We used the same approach to examine how often the alternative patch with the highest replenishment rate was selected, except that we extracted replenishment rates for the two alternative patches and computed the proportion of choices in which the highest replenishment patch was selected.

***Average Reward Rates***

To test the prediction that the average reward rate would be higher under free choice conditions, we extracted the number of points earned in forced and free choice blocks. These values were then divided by the total number of actions in each condition (i.e. 400). Once this procedure had been performed for each participant, the average reward rates for free and forced choice conditions were compared across participants using a paired two-tailed t-test. To examine the average reward rate per exploit action, we divided the number of points earned in each condition by the number of exploit actions made in that condition. The reward rates per exploit action for forced and free choice conditions were compared using a paired two-tailed t-test. To explore this effect in more depth, the reward rate per exploit action was compared between conditions, separately for each patch (slow, medium, fast replenishment rate) with paired two-tailed t-tests. The alpha threshold for the three exploratory tests was corrected using the Bonferroni correction.

***Reward Thresholds for Leaving***

To test whether participant's reward thresholds for leaving the current patch were different between free and forced choice conditions, we extracted the last reward outcome prior to each leave decision. We then averaged the reward outcome before leaving, separately for each choice condition. Once this procedure had been performed for all participants, a 2 x 3 repeated measures ANOVA was performed on the reward outcome before leaving, with factors of choice condition (free x forced) and patch (slow, medium, fast replenishment rate). The main effect of patch was followed up with paired t-tests that compared the mean reward before leaving between the fast and medium replenishing patches, the fast and slow replenishing patches, as well as the medium and slow replenishing patches. The alpha threshold was corrected for three follow up tests using the Bonferroni correction.

***Rewards on Arrival***

To test whether rewards were higher when participants arrived at patches, we extracted the number of coins earned for the first exploit action, each time the slow, medium and fast replenishing patches were visited. We then averaged the arrival rewards, for each patch and each condition. We then performed a 2x3 repeated measures ANOVA on the arrival rewards, with factors of choice condition (free x forced) and the arrival patch (slow, medium, fast replenishment rate). The interaction was followed up with paired t-tests that compared the arrival reward between choice conditions, separately for each arrival patch. The alpha threshold for the three exploratory tests was corrected using the Bonferroni correction. A similar procedure was performed to examine the average arrival reward, as a function of the patch being left. The critical difference was that we extracted the first reward when arriving at the next patch (regardless of its replenishment rate), each time participants decided to leave the slow, medium or fast replenishing patch. The main effect of the patch being abandoned was followed up with separate paired t-tests that compared mean arrival rewards between specific patches (i.e. slow vs medium, slow vs fast, medium vs fast). The alpha threshold was therefore corrected for three exploratory tests using the Bonferroni correction

***Actions Before Leaving***

To examine the number of actions prior to leaving patches, we extracted the number of exploit actions before each leave decision, in the forced and free choice blocks. We then averaged the exploit actions, separately for each condition. Once this procedure had been

performed for all participants, a 2 x 3 repeated measures ANOVA was performed on the number of exploit actions before leaving, with factors of choice condition (free x forced) and patch (slow, medium, fast replenishment rate). In addition, we performed exploratory analyses that examined whether leaving times differed between choice conditions for each patch individually. Exploratory analyses followed the same procedure above, except that the number of exploit actions prior to leaving were averaged separately for each patch and compared between choice conditions using separate paired two-tailed t-tests. The alpha threshold for the three exploratory tests was corrected using the Bonferroni correction.

***Use of Global vs Local Reward for Leave Decisions***

Our main analysis concerned the comparison of the local and global model described in the main text. The four free parameters for each model were fit for each participant using a scatter-based optimisation solver in Matlab. For all models, the constant, $c$, was constrained between -50 and +50, the parameter $\beta$ was constrained between 0 and +2 and learning rates (one for updating estimates after reward outcomes and one for down-weighting) were constrained between 0 and +1. Models were fit to exploit-or-leave choices, separately for free and forced conditions. For each trial in a condition, the exploit probability was calculated using the logistic decision equation reported in the main text. One exception to this was on the first decision when arriving at a patch, where participants were forced to exploit and the exploit probability was set to 1-(1e-5), without using the logistic choice equation. On all trials, exploit probabilities were constrained to a maximum value of 1-(1e-5) and a minimum value of 1e-5. The leave probability was calculated as 1-p(exploit). The probability of the choice made on the current trial was stored. The negative log of the choice probabilities were summed to get the negative log likelihood of the model given the data. Model parameters were selected on the basis of minimising the negative log likelihood. The performance of each model was then compared using the Bayesian Information Criterion (BIC; Schwarz, 1978). The BIC was calculated for each model, for each participant. The model with the lowest BIC across all participants was used as the reference model. For each participant, the BIC score for this reference model was subtracted from the BIC scores of the two remaining models. Evidence for one model over another was assessed using bayes factors, computed based on the difference in BIC scores, using a Matlab toolbox (https://klabhub.github.io/bayesFactor/).

**Author contributions**

Conceptualisation, S.H., P.D., N.W.S; Methodology, S.H., P.D., N.W.S; Investigation, S.H.; Writing - Original Draft, S.H; Writing - Review and Editing, S.H., P.D., N.W.S; Visualisation, S.H, N.W.S.; Supervision, N.W.S; Funding Acquisition, N.W.S.

**Acknowledgments**

The authors thank Nir Moneta for technical advice and Noa L. Hedrich for comments on the manuscript. This research was funded by an Independent Max Planck Research Group grant from the Max Planck Society (M.TN.A.BILD0004) and a Starting Grant from the European Union (ERC-2019-StG REPLAY-852669) awarded to N.W. Schuck. PD was funded by the Max Planck Society and the Alexander von Humboldt Foundation.

**Declaration of interests**

The authors declare no competing interests.

## References


Barack, D. L., Chang, S. W. C., & Platt, M. L. (2017). Posterior cingulate neurons dynamically signal decisions to disengage during foraging. *Neuron*, *96*(2), 339–347.

Charnov, E. L. (1976). Optimal foraging, the Marginal Value Theorem. *Theoretical Population Biology*, *9*, 129–136.

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.

Erwin, R. M. (1985). Foraging decisions, patch use, and seasonality in egrets (Aves: Ciconiiformes). *Ecology*, *66*(3), 837-844.

Fouragnan, E. F., Chau, B. K., Folloni, D., Kolling, N., Verhagen, L., Klein-Flügge, M., ... & Rushworth, M. F. (2019). The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. *Nature Neuroscience*, *22*(5), 797-808.

Garcia, L. C., & Eubanks, M. D. (2019). Overcompensation for insect herbivory: A review and meta-analysis of the evidence. *Ecology*, *100*(3), e02585.

Hall-McMaster, S., & Luyckx, F. (2019). Revisiting foraging approaches in neuroscience. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(2), 225-230.

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, *14*, 933–939.

Huys, Q. J., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol*, *8*(3), e1002410.

Huys, Q. J., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., ... & Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, *112*(10), 3098-3103.

Kamil, A. C. (1978). Systematic Foraging by a Nectar-Feeding Bird, the Amakihi (Loxops virens). *Journal of Comparative and Physiological Psychology*, *92*(3), 388-396.

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 and Behavioral Neuroscience*, *17*(6), 1–11.

Kilpatrick, Z. P., Davidson, J. D., & Hady, A. E. (2020). Normative theory of patch foraging decisions. *arXiv preprint arXiv:2004.10671*.

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural mechanisms of foraging. *Science*, *336*(6077), 95–98.

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.

Kurth-Nelson, Z., Economides, M., Dolan, R. J., & Dayan, P. (2016). Fast sequences of non-spatial state representations in humans. *Neuron*, *91*(1), 194-204.

Le Heron, C., Kolling, N., Plant, O., Kienast, A., Janska, R., Ang, Y. S., ... & Apps, M. A. (2020). Dopamine modulates dynamic decision-making during foraging. *Journal of Neuroscience*.

Lihoreau, M., Chittka, L., & Raine, N. E. (2010). Travel optimization by foraging bumblebees through readjustments of traplines after discovery of new feeding locations. *The American Naturalist*, *176*(6), 744-757.

Lottem, E., Banerjee, D., Vertechi, P., Sarra, D., Lohuis, M. O., & Mainen, Z. F. (2018). Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nature Communications*, *9*(1), 1–12.

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.

Merkle, J. A., Potts, J. R., & Fortin, D. (2017). Energy benefits and emergent space use patterns of an empirically parameterized model of memory-based patch selection. *Oikos*, *126*(2), 185–195.

Mobbs, D., Trimmer, P. C., Blumstein, D. T., & Dayan, P. (2018). Foraging for foundations in decision neuroscience: Insights from ethology. *Nature Reviews Neuroscience*, *19*, 419–427.

Passingham, R. E. (1985). Memory of monkeys (Macaco mulatta) with lesions in prefrontal cortex. *Behavioral Neuroscience*, *99*(1), 3–21.

Possingham, H. P., & Houston, A. I. (1990). Optimal patch use by a territorial forager. *Journal of Theoretical Biology*, *145*(3), 343–353.

Riotte-Lambert, L., Benhamou, S., & Chamaillé-Jammes, S. (2015). How memory-based movement leads to nonterritorial spatial segregation. *The American Naturalist*, *185*(4), E103–E116.

Rushworth, M. F., 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.

Sayers, K., & Menzel, C. R. (2012). Memory and foraging theory: Chimpanzee utilization of optimality heuristics in the rank-order recovery of hidden foods. *Animal Behaviour*, *84*(4), 795–803.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.

Seidel, D. P., & Boyce, M. S. (2015). Patch-use dynamics by a large herbivore. *Movement Ecology*, *3*(1), 7.

Stephens, D. W., & Krebs, J. R. (1986). Foraging theory. Princeton: Princeton University Press.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction. Cambridge: MIT press.

Sweis, B. M., Thomas, M. J., & Redish, A. D. (2018). Mice learn to avoid regret. *PLoS Biology*, *16*(6), e2005853.

Wittmann, M. K., Kolling, N., Akaishi, R., Chau, B. K., Brown, J. W., Nelissen, N., & Rushworth, M. F. (2016). Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. *Nature Communications*, *7*(1), 1-13.