Classification: Biological Sciences: Psychology and Cognitive Sciences

Rats exhibit similar biases in foraging and intertemporal choice tasks

Gary A. Kane[1,2], Aaron M. Bornstein[1], Amitai Shenhav[3], Robert C. Wilson[4], Nathaniel D. Daw[1], Jonathan D. Cohen[1]

[1]Department of Psychology and Princeton Neuroscience Institute, Princeton University, Washington Rd., Princeton, NJ, 08544
[2]The Rowland Institute at Harvard, Harvard University, 100 Edwin H. Land Blvd., Cambridge, MA, 02142
[3]Department of Cognitive, Linguistic, and Psychological Sciences and Carney Institute for Brain Science, Brown University, 190 Thayer St., Providence, RI, 02912
[4]Department of Psychology and Cognitive Science Program, University of Arizona, 1503 E University Blvd., Tucson, AZ, 85721

Corresponding author: Gary A. Kane, gkane@rowland.harvard.edu

Keywords: intertemporal choice, decision-making, foraging, delay discounting

Abstract

In patch foraging tasks, animals consistently overharvest — they stay in a patch with diminishing rewards longer than is predicted by optimal foraging theory. While overharvesting appears to be ubiquitous, its cause remains poorly understood. To address this question, we characterized the behavior of rats in a series of patch foraging tasks. Most notably, rats left patches earlier if a delay was imposed between the decision to harvest and receiving reward (pre-reward delay), and rats were sensitive to increasing the delay between receiving reward but before the next decision to harvest or leave (post-reward delay). This behavior was best explained by a temporal discounting model that maximizes all future reward. To test the external validity of this hypothesis, rats were further tested in a standard delay discounting task, in which we separately manipulated pre- vs. post-reward delays. As in the foraging paradigm, rats were less likely to select larger rewards if the pre-reward delay was longer and rats were sensitive to post reward delays. Strikingly, parameters obtained from fitting the temporal discounting model to foraging behavior provided an excellent fit to behavior in the delay discounting task. These results suggest that rats employ a common computational approach to both foraging and delay-discounting decisions, and that overharvesting may reflect operation of the same mechanisms that underlie discounting of future reward.

Significance Statement

Two common paradigms to study decision making include foraging tasks, in which animals choose to accept an available reward vs. reject it to search for better opportunities, and intertemporal choice tasks, in which animals choose between a smaller reward after a short delay or larger reward after a longer delay. In both tasks, animals exhibit a preference for the more immediate reward. Interestingly, a common model of intertemporal choice behavior — temporal discounting — has failed to explain the similar bias in foraging tasks. Using carefully designed behavioral experiments and quantitative analysis, we show that rats exhibit identical time preferences in the two tasks, and their behavior is well explained by a temporal discounting model that captures important features of both tasks.

Introduction

In foraging paradigms, animals must choose whether to continue to harvest depleting reward within a patch, or incur a cost of time and effort to travel to a new patch that is expected to have a higher value. Animals qualitatively follow predictions of the optimal behavior described by the Marginal Value Theorem (MVT) (1) — to leave a patch when it depletes to the average reward rate across all patches. Consistent with MVT, animals leave a patches earlier if they yield less reward than average or if the cost of traveling to a new patch is greater. However, animals consistently show a bias toward overharvesting; that is, they continue to harvest from a patch beyond the point at which it depletes to the average reward rate (2–5). Previous hypotheses to explain overharvesting include subjective costs, such as an aversion to leaving a patch (6, 7), and diminishing marginal utility, by which larger rewards in a new patch are not perceived as proportionally larger than smaller depleted rewards in the current patch (2). But these hypotheses have never been systematically compared in a set of experiments designed to directly test their predictions. Furthermore, these hypotheses are not compatible with the widely observed preference for smaller, more immediate rewards compared to larger, delayed ones in other intertemporal choice tasks.

Preference for small, immediate rewards over larger, delayed rewards in intertemporal choice tasks (also referred to as self-control or delay discounting tasks) (8, 9), in which animals are presented with two options simultaneously, are commonly explained via two models: temporal discounting and short-term rate maximization. Temporal discounting models predict that animals discount the value of a reward by the time it takes to receive it. Discounting can be adaptive in unstable environments — if the environment is likely to change before future rewards can be acquired, it is appropriate to place greater value on rewards available in the near future. Under this hypothesis, the normative discount function is exponential — the rate of discounting remains constant over time (10, 11). However, animal preferences typically follow a hyperbolic form — the rate of discounting is steeper initially and decreases over time (10, 11). This yields inconsistent time preferences or preference reversals: those who are inclined to not wait for a large reward after longer delay may change their mind and prefer to wait for the larger reward as the time to receive rewards draws near (8–11).

Short-term maximization rules predict that animals seek to maximize reward rate locally (i.e. the value of the next reward over the time to receive it), ignoring opportunity costs of the decision at hand (any time delays after receiving reward) (12–16). One advantage to maximizing local reward rates is that animals may be able to better estimate the value of rewards in the short-term, thus making better decisions over this time than if they considered all future reward (14, 15). In support of this hypothesis, in intertemporal choice tasks, animals typically attend only to delays between decisions and receiving rewards and they are insensitive to post-reward delays i.e., delays between receiving reward and making the next decision (12, 13, 17, 18).

In the present study, we sought to determine the biases that underlie suboptimal decision making in foraging tasks. We tested rats in a series of patch foraging tasks designed to test the predictions of the hypotheses described above. In certain environments, rats deviated from predictions of MVT: despite equivalent reward rates, rats stayed longer in patches when given larger rewards over longer time delays and rats left patches earlier when they had to wait to receive reward after deciding to harvest — a bias that is strikingly similar to delay discounting observed in intertemporal choice tasks. The latter finding could not by subjective costs and diminshing marginal utility, but could be explained by temporal discouting and short-term rate maximization. We further tested rats time preferences by examining sensitivity to post-reward delays in both

foraging and a standard intertemporal choice task. Contrary to previous studies (12, 13, 17, 18), rats were sensitive to post-reward delays in both paradigms. To understand the mechanism driving these decision biases, we used quantitative model comparison to test predictions of several possible accounts for this pattern of behavior. A form of hyperbolic discounting provided the best explanation for rat behavior across all experiments. These data suggest that rats exhibit similar biases in foraging and two-alternative choice paradigms, and maximization of discounted future rewards may be a common mechanism for suboptimal decision-making across paradigms.

Results

Rats consider long-term rewards, but exhibit a bias in processing pre- vs. post-reward delays

Long Evans rats (n = 8) were tested in a series of patch foraging tasks in operant conditions chambers (19). To harvest reward (10% sucrose water) from a patch, rats pressed a lever at the front of the chamber, and reward was delivered in an adjacent port. After a post-reward delay (inter-trial interval or ITI), rats again chose to harvest a smaller reward, or to leave the patch by nose poking in the back of the chamber. A nose poke to leave the patch caused the harvest lever to retract and initiated a delay simulating the time to travel to the next patch. After the delay, the opposite lever extended, and rats could then harvest from (or leave) this replenished patch (Fig. S1). In four separate experiments, we manipulated different variables of the foraging environment: i) a 10 s vs. 30 s travel time between patches, ii) reward depletion rate of 8 vs. 16 µL, iii) the magnitude of rewards and delays, such that in one condition, the size of rewards and length of delays was twice that of the other (“scale“ experiment), and iv) the placement of delays (“handling time“ experiment): the total time to harvest reward remained constant, but in one condition there was no pre-reward delay and ~13 s post reward delay, and in the other there was a 3 s handling time (pre-reward delay) simulating the time obtain reward once deciding to harvest, and ~10 s post-reward delay. Parameters for each experiment are shown in Table 1.

Experiment	Condition	Start Reward	Depletion Rate	Pre-Reward Delay	Harvest Time	Travel Time
travel time	10s 30 s	60, 90, or 120 µL	-8 µL	0 s	10 s	10 s 30 s
depletion rate	-8 µL -16 µL	90 µL	-8 µL -16 µL	0 s	12 s	12 s
scale	90 µL/10 s 180 µL/20 s	90 µL 180 µL	-8 µL -16 µL	0 s	10 s 20 s	10 s 20 s
handling time	0 s 3 s	90 µL	-8 µL	0 s 3 s	15 s	15 s

Table 1. Parameters for each of the first 4 foraging experiments. Harvest time = time to make a decision to harvest + pre-reward delay + inter-trial interval. To control reward rate in the patch, the inter-trial interval was adjusted relative to the decision time to hold the harvest time constant.

The first experiment (“travel time“) was designed to test the two main predictions of MVT: i) that animals should stay longer in patches that yield greater rewards and ii) animals should stay longer in all patches when the cost of traveling to a new patch is greater. In this experiment, rats

encountered three different patch types within sessions, which started with varying amount of reward (60, 90, or 120 µL) and depleted by the same rate (8 µL/trial). The delay between patches was either 10 s or 30 s; each travel time delay was tested in its own block of sessions and the order of these was counterbalanced across rats. As predicted by MVT, rats stayed longer in patch types that started with larger reward volume, indicated by more harvests per patch (β = 118.266, SE = 2.753, $p < .001$). Rats also stayed longer in all patch types when time between patches was longer (β = 1.907, SE = .306, $p < .001$; Fig. 1A). However, rats uniformly overharvested relative to predictions of MVT ($\beta_{\text{rat-MVT}}$ = 3.396 trials, SE = .176, $p < .001$). The degree to which rats overharvested was not significantly different between the 10 s and 30 s travel conditions ($\beta_{\text{10 s-30 s}}$ = .304 trials, SE = .155, p = .088).

The second experiment ("depletion rate") tested another central variable in foraging environments: the rate of reward depletion within a patch. Quicker reward depletion causes the local reward rate to deplete to the long-run average reward rate quicker, thus MVT predicts earlier patch leaving. Within sessions, rats encountered a single patch type (starting volume of 90 µL) that depleted at a rate of either 8 or 16 µL/trial, tested in separate sessions and counterbalanced. As predicted by MVT, rats left patches earlier when patches depleted more quickly (β = 2.589, SE = .155, $p < .001$; Fig. 1B). But, again, rats stayed in patches longer than is predicted by MVT ($\beta_{\text{rat-MVT}}$ = 2.000 trials, SE = .134, $p < .001$). Rats overharvested to a greater degree in the 8 µL depletion condition than the 16 µL depletion condition ($\beta_{\text{8 µL-16 µL}}$ = 1.589 trials, SE = .155, $p < .001$).

Because MVT is concerned with reward-rate optimization, MVT predicts that a manipulation that increases reward size in proportion to reward delay should have no effect on the number of harvests in a patch. We tested this prediction in a third experiment ("scale"). We manipulated the scale of rewards and delays in the following manner: patches started with (A) 90 or (B) 180 µL of reward, depleted at a rate of (A) 8 or (B) 16 µL/trial, and the duration of harvest trials and travel time between patches was (A) 10 or (B) 20 s. Rats overharvested in both A and B conditions ($\beta_{\text{rat-MVT}}$ = 4.374 trials, SE = .153, $p < .001$) and, contrary to predictions of MVT, rats stayed in patches significantly longer in the B condition that provided larger rewards but at proportionately longer delays (β = 1.937, SE = .193, $p < .001$; Fig. 1C).

In delay discounting tasks, animals are more sensitive to pre-reward delays than post-reward delays — they value a reward less when it is received after a delay, but are typically not affected by delays between receiving reward and the start of the subsequent trial. In a final behavioral experiment (handling time) with this cohort of rats, we directly tested sensitivity to pre- vs. post-reward delays in the context of foraging. In one condition rats received reward immediately after lever press followed by a post-reward delay of ~13 s before the start of the next trial. In the other condition, there was a 3 s pre-reward delay between lever press and receiving reward followed by a shorter post-reward delay of ~10 s. The total time of each trial was held constant between conditions; there was no difference in reward rate and therefore MVT predicts no change in number of harvests between conditions. Contrary to this prediction, rats left patches earlier in the environment with a 3 s pre-reward delay and shorter post-reward delay (β = 2.345, SE = .313, $p < .001$; Fig. 1D). This result suggests that rats value reward within the patch less when they have to wait to receive it.

To better understand whether rats are hypersensitive to pre-reward delays or insensitive to post-reward delays, a new cohort of rats (n = 8) was tested for their sensitivity to post-reward delays in a foraging task, and separately for their sensitivity to both pre- and post-reward delays in a standard intertemporal choice task. In this foraging experiment, rats were tested in two conditions: one with a short post-reward delay (3 s) and the other with a longer post-reward delay

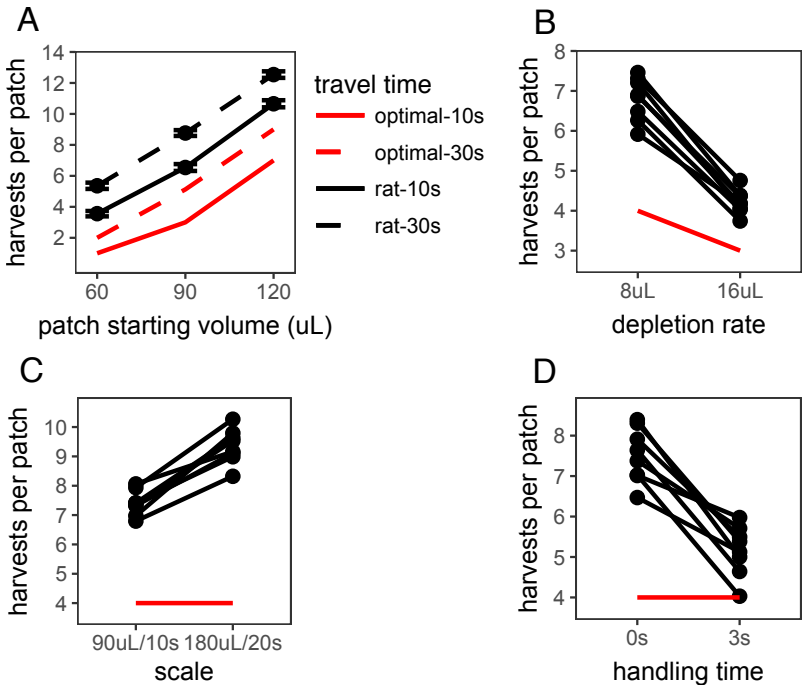

Figure 1. Rat foraging behavior in the A) travel time, B) depletion rate, C) scale, and D) handling time (pre-reward delay) experiments. In A, points and error bars represent mean ± standard error. In B-D, points and connecting lines represent behavior of each individual rat. Red lines indicate optimal behavior (per MVT).

(12 s). The total time of harvest trials was not held constant; the longer post-reward delay reduced reward rate within patches, such that MVT predicts that rats should leave patches earlier. In contrast to predictions of MVT, prior studies of standard intertemporal choice behavior suggest that animals are insensitive to post-reward delays, indicating that they are only concerned with maximizing short-term reward rate (the time from their response to the associated reward) (13, 18). This hypothesis predicts that the increase in post-reward delay should have no effect on rat behavior. As predicted by MVT and not short-term rate maximization, rats were sensitive to the post-reward delay, leaving patches earlier in the 12 s delay condition (β = 1.411, SE = .254, $p < .001$; Fig. 2A). As in other experiments, rats overharvested ($\beta_{\text{rat-MVT}}$ = 3.332 trials, SE = .285, $p < .001$), and there was no difference in the degree to which rats overharvested between the 3 s and 12 s delay conditions ($\beta_{3\text{ s-}12\text{ s}}$ = .340 trials, SE = .286, p = .274).

We next examined whether the same rats would exhibit similar time preferences (hypersensitivity to pre-reward delays while remaining sensitive to post-reward delays) in a delay-discounting task. On each trial, rats pressed either the left or right lever to receive a smaller-sooner reward of 40 µL after a 1 s delay or a larger-later reward of 40, 80, or 120 µL after a 1, 2, 4, or 6 s delay. The task consisted of a series of 20 trial episodes. At the start of the episode, the larger-later reward value and delay, and larger-later lever (left or right) were randomly selected. For the first 10 trials of each game, rats were forced to press either the left or right lever to learn the value and delay associated with that lever (only one lever extended on each of these trials). For the last 10 trials, both levers extended and rats were free to choose. Rats were tested on two different versions of this task: one in which the post-reward delay was held constant, such that the longer pre-reward delays reduced reward rate (constant delay); and another in which the time of the trial was held constant, such that longer pre-reward delays resulted in shorter post-reward

delays to keep reward rate constant (constant trial). MVT, which maximizes long-term reward rate, predicts that rats would be sensitive to the pre-reward delay in the constant delay condition, but not the constant trial condition (in which the pre-reward delay does not affect reward rate).

Rats exhibited a change in choice preference over the course of the 10 free choice trials, indicating that they might have still been learning reward values in early free choice trials. Accordingly, we focused analysis on the final 5 free choice trials (statistical analyses were robust to including all 10 free choice trials). In both conditions, rats were sensitive to pre-reward delays — they were less likely to select the larger-later reward with longer pre-reward delays (Fig. 2B; effect of delay within conditions: $\chi^2(3) = 28.633$, $p_{\text{const delay}} < .001$; $\chi^2(3) = 12.946$ $p_{\text{const trial}} = .004$). However, rats were less sensitive to the pre-reward delay in the constant trial condition, in which it did not affect reward rate (effect of delay between conditions: $\chi^2(3) = 9.437$, $p = .024$). This effect was most evident at the longest pre-reward delay (Fig. 2C; $\chi^2(1) = 6.453$, $p_{\text{const delay:6s vs. const time:6s}} = .044$).

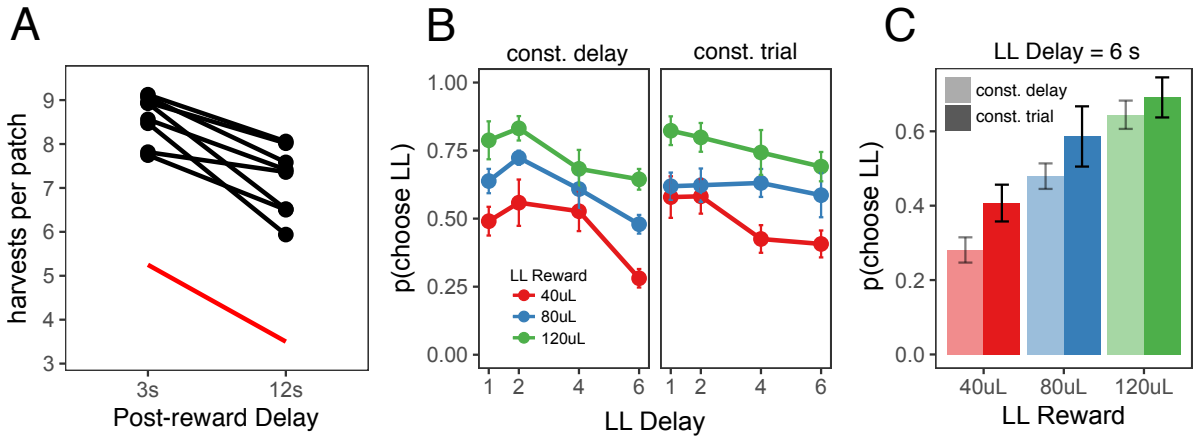

Figure 2. A) Rat behavior in the post-reward delay experiment. Points and lines represent behavior of individual rats. Red line indicates optimal behavior (per MVT). B) Rat behavior in the two-alternative intertemporal choice task. Points and error bars represent mean $\pm$ standard error for each condition. C) Comparison of behavior across the constant delay and constant trial versions of the two-alternative choice task with the 6 s delay (same data as in B). Error bars represent standard error.

Quasi-hyperbolic discounting best explains behavior across all tasks

Differential sensitivity to time delays that precede rewards (while anticipating reward delivery) vs. those that follow reward receipt (before making a subsequent decision) suggests a preference for rewards in the near future. To formally test specific hypotheses about the basis for this behavior, we modeled both tasks as continuous time semi-markov processes. The time between each event in the task was represented as a state (e.g. cues turning on/off, lever press, reward delivery; state space diagram of foraging task in Fig. S2). The value of a state was defined as the discounted value of all future rewards available from that state. As the discount factor approached 1 (i.e. no temporal discounting), this model converged to long-term reward maximization, equivalent to MVT. Full details for all models can be found in the methods and supporting information.

Subjective costs to leave a patch and diminishing marginal utility for larger rewards have both been previously hypothesized as causes for overharvesting, and both have been used to explain overharvesting in a limited number of studies (2, 6, 7). As these hypotheses are based on MVT, they are only concerned with reward rate, and are indifferent to the placement of delays. Thus, they cannot explain the time preferences exhibited by the rats in our foraging task. To confirm these hypotheses would fail to explain rat behavior in the present experiments, we implemented a model for each of these hypotheses and fit them to rat foraging behavior. We found that these hypotheses could explain overharvesting exhibited in the patch starting reward and depletion rate experiments, but not time preferences in the handling time experiment (Fig. S3; for more details, see supporting information).

We focused on two classes of models that assume biased perception of reward value over time: ignoring (or underestimating) post-reward delays and temporal discounting. Models in the first class, which assumed greater reward sensitivity to pre-reward delays than post-reward delays, implement forms of short-term maximization rules. We tested three models of this type, in which: A) post-reward delays are perceived as shorter than they actually are in a linear fashion; B) post-reward delays are perceived as shorter than they actually are in a non-linear fashion, in which sensitivity to the delay decreases as the delay increases; and C) pre-reward delays are perceived as longer than they actually are in a linear fashion. Model C represented an aversion to time spent in anticipation of receiving reward. Model B, that implemented decreasing sensitivity to the post-reward delay, provided the best fit to the first four experiments (Fig 3A-D). However, this model largely ignored longer post-reward delays, and could not account for rats sensitivity to post-reward delays (Fig 3E). Model A, linear underestimation of post-reward delays, and Model C, overestimation of pre-reward delays, predicted the direction of all effects in all experiments: overharvesting in the travel time and depletion rate experiments, staying longer in patches due to larger rewards with longer delays, earlier patch leaving due to a pre-reward delay, and sensitivity to post-reward delays. However, these models failed to predict the magnitude of the changes in the scale and handling time experiments (Fig 3C-D).

We also tested three models in the second class, that explicitly implemented temporal discounting: an exponential, a hyperbolic, and a quasi-hyperbolic discounting model. To implement the standard form of hyperbolic discounting in our semi-markov model, we used an approach previously described by Kurth-Nelson & Redish, 2009 (20). This model contained a number of “µAgents,“ each with their own exponential discount factor and value function. The overall value function of the “macroAgent“ was the average of the µAgents. The rate of hyperbolic discounting exhibited by the macroAgent is determined by the distribution of exponential discount factors of the µAgents. The quasi-hyperbolic discounting model was a more flexible form of hyperbolic discounting, allowing for steeper discounting of rewards obtained in the near future relative to rewards obtained further in the future (21). This model consisted of a weighted-sum of two exponentially discounted values (22). All three discounting models explained overharvesting in the travel time and depletion rate experiments, and that rats would stay longer in patches yielding larger rewards with longer delays (Fig 4A-C). Surprisingly, the exponential and hyperbolic discounting models did not predict that rats would leave earlier due to a pre-reward delay, but the quasi-hyperbolic discount model did explain this bias. As these models considered all future rewards, all three models predicted rats would be sensitive to the post-reward delay (Fig 4E).

The quasi-hyperbolic discounting model was the only model tested that could qualitatively capture rat behavior across all foraging tasks. To determine which model provided the best quantitative fit, we examined the group-level Bayes Information Criterion (integrated BIC or iBIC) (23, 24) of all models in each of the foraging tasks (Fig. 5, S4). The quasi-hyperbolic discount-

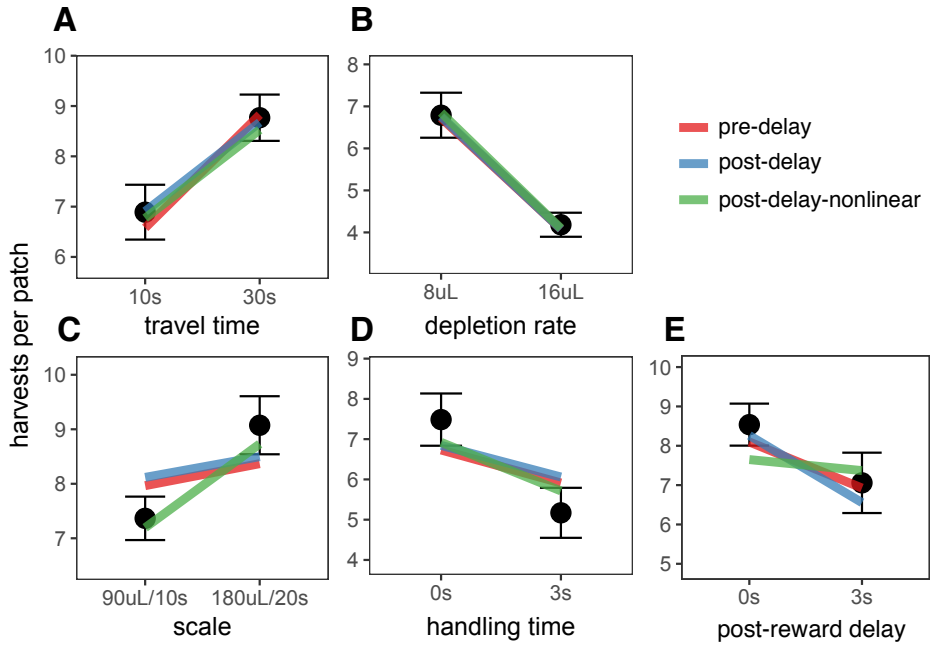

Figure 3. Predictions of the best fit models of overestimation of pre-reward delays (pre-delay), linear underestimation of post-reward delays (post-delay), and nonlinear underestimation of post-reward delays (post-delay-sublinear). Points and errorbars are the mean $\pm$ standard deviation of rat behavior, colored lines represent the mean predicted number of harvests across all rats.

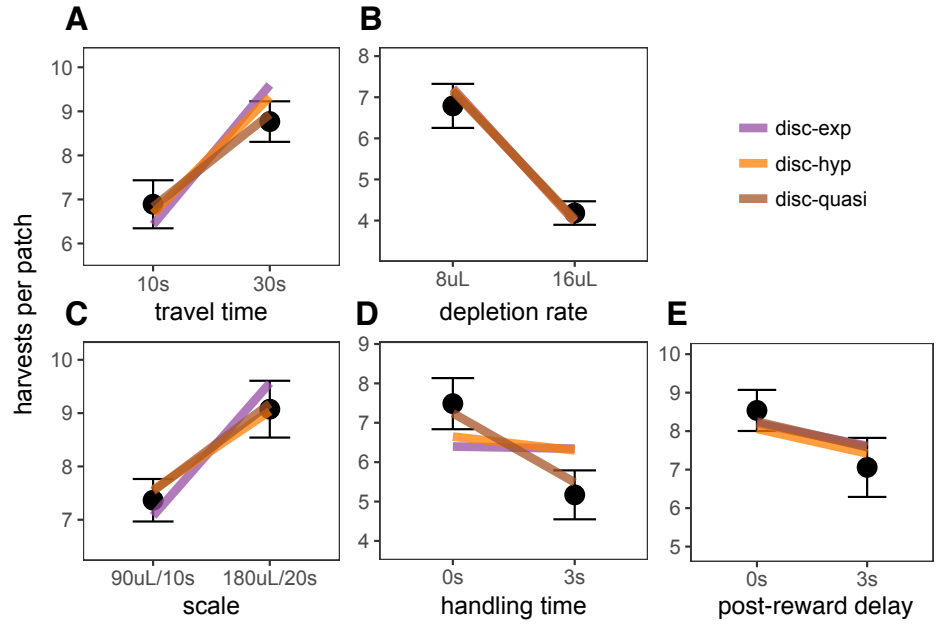

Figure 4. Predictions of the best fit exponential discounting model (disc-exp), hyperbolic discounting model (disc-hyp), and quasi hyperbolic discounting model (disc-quasi). Points and error bars are the mean $\pm$ standard deviation of rat behavior; colored lines represent the mean predicted number of harvests across all rats.

ing model had the lowest cumulative score over all foraging tasks (Fig. 5A) and also scored the lowest in iBIC in all individual foraging experiments except for the depletion rate experiment, in which it ranked second to the hyperbolic discounting model with a difference of 6 ($iBIC_{hyperbolic}$ = 5726.858, $iBIC_{quasi\text{-}hyperbolic}$ = 5732.990; Fig. S4).

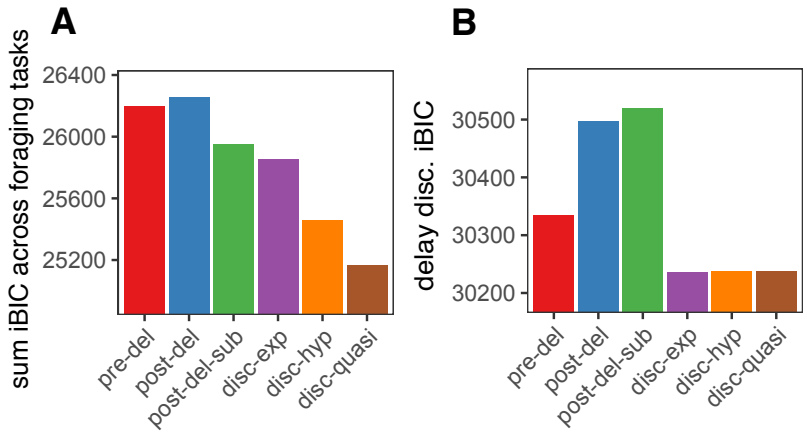

Figure 5. A) Sum of iBIC scores for each model across all 5 foraging experiments. B) iBIC score for each model for the delay discounting experiment, using only the last 5 choices of each episode. Pre-del = linear overestimation of pre-reward delays, post-del = linear underestimation of post-reward delays, post-del-non = nonlinear underestimation of post-reward delays, disc-exp = exponential discounting, disc-hyp = hyperbolic discounting, disc-quasi = quasi-hyperbolic discounting.

We next determined whether the same quasi-hyperbolic discounting model could explain behavior in the delay discounting task. For the delay discounting task, each series of 10 free choices was modeled as a separate episode, such that the value of the first choice was equal to discounted reward across all 10 choices, the value of the second choice equal to discounted reward across the remaining 9 choices, and so on (see abbreviated state space diagram in Fig. S5A). The episode ended at the 10th choice, and did not consider rewards on future games. Consistent with prior analyses, the model was fit to only the last five choices of each episode. We tested both classes of models (underestimating post-reward delays and temporal discounting) in this task, and found that all three temporal discounting models had lower iBIC scores than all three of the underestimating post-reward delay models. The exponential discounting model had the lowest iBIC, but differences in iBIC score among the three temporal discounting models were minor ($iBIC_{exponential}$ = 30234.9, $iBIC_{hyperbolic}$ = 30237.61, $iBIC_{quasi\text{-}hyperbolic}$ = 30236.94). Similar results were found when fitting the model to all 10 free choices (Fig. S5C).

To further test whether quasi-hyperbolic discounting may reflect the operation of a common mechanism for suboptimal decision-making across tasks, we tested whether the quasi-hyperbolic discounting model fit to one task could predict behavior of the same rats in the other task. Data from both tasks were separated into thirds. The quasi-hyperbolic discounting model was fit to two-thirds of the data from one task, and we calculated the negative log marginal likelihood across all rats (-LL) (23, 24) for the remaining one-third of data from both tasks. This procedure was repeated such that each third of the data served as the left-out sample, and we took the sum of the negative log marginal likelihood across each left-out sample. Surprisingly, parameters fit to the foraging task provided a better fit to the left out sample of data in both the foraging and intertemporal choice tasks (foraging task: $-LL_{forage\ pars}$ = 5166, $-LL_{DD\ pars}$ = 6351;

ITC task: $-LL_{\text{forage pars}} = 31343$, $-LL_{\text{DD pars}} = 32469$; Fig 6). These data suggest that overharvesting in foraging tasks and time preferences in intertemporal choice tasks can be explained via a common mechanism.

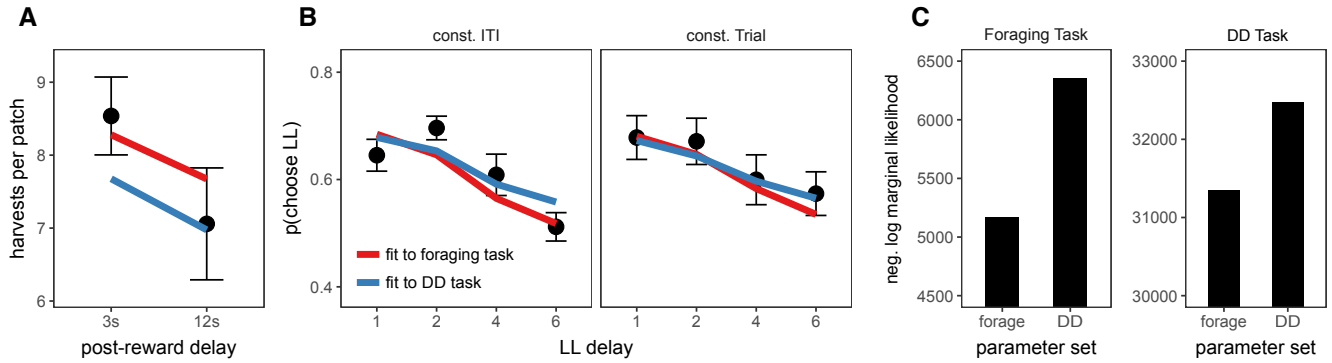

Figure 6. A) Predicted foraging behavior for quasi-hyperbolic model parameters fit to either the foraging task (red line) or delay discounting task (DD; blue line). Black points and error bars represent rat data. B) Predicted delay discounting behavior for quasi-hyperbolic model parameters fit to data from either the foraging or delay discounting task, plotted against rat behavior. C) Negative log marginal likelihood of the left out sample of foraging data (left) or delay discounting data (right) for parameters fit to each task.

Discussion

In foraging studies, animals exhibit behavior that conforms qualitatively to predictions made by optimal foraging theory (i.e., the MVT), choosing to leave a patch when its value falls below that of the average expected value of other(s) available in the environment. However, an almost ubiquitous finding is that they overharvest, leaving a patch when its value falls to a value lower than the one predicted by MVT. Given that the rewards available within the current patch are generally available sooner than those at other patches due to travel time, one interpretation of overharvesting is that this reflects a similarly prevalent bias observed in intertemporal choice tasks, in which animals consistently show a greater preference for smaller more immediate rewards than later delayed rewards that would be predicted by optimal (i.e,. exponential) discounting of future values. However, in prior studies, models of intertemporal choice behavior have been poor predictors of foraging behavior (7, 25). Here, we show that in a carefully designed series of experiments, rats exhibit similar time preferences in foraging and intertemporal choice tasks, and that at least one model of intertemporal choice, one that assumes quasi-hyperbolic discounting, can explain the rich pattern of behaviors observed in both foraging and an intertemporal choice tasks.

The foraging behavior we observed was consistent with that observed in previous studies of foraging behavior in rats, monkeys, and humans (2, 3, 19). Specifically, rats followed qualitative but not quantitative predictions of MVT: they stayed longer in patches that yielded greater rewards, they stayed longer in all patch types when the cost of traveling to a new patch was greater, and left patches earlier when rewards depleted more quickly; but they also consistently overharvested (they stayed longer in patches than is predicted by MVT). Our experiments also revealed novel aspects of overharvesting behavior, demonstrating that in certain environments

rats qualitatively violate MVT. Specifically, we showed that rats overharvested more when reward amount and delay were increased, even though reward rate was held constant, and that rats were differentially sensitive to whether the delay was before the receipt of the proximal reward, or following reward delivery. These findings supported the conjecture that overharvesting is related to time preferences. When we directly tested rats sensitivity to post-reward delays, we found that contrary to previous studies addressing time preferences (12, 13, 18, 26), rats were sensitive to post-reward delays in both the foraging task and the delay discounting task.

Whether animals maximize long-term rewards (i.e., the reward rate across all patches or all future decisions) has long been under debate (12, 27, 28). Impulsivity observed in delay discounting tasks (e.g., two-alternative intertemporal choice tasks) suggests that animals may maximize local rather than longterm reward rate, failing to fully consider opportunity costs of that decision (such as a long post-reward delay) (5, 10, 12, 28, 29). This is similar to time preferences expressed in standard delay discounting models (10, 12). In contrast, findings from the present study provide at least two lines of evidence that rats consider opportunity costs in intertemporal choice: i) in an environment with multiple patch types (travel time experiment), they stayed longer in patches that yielded greater rewards (i.e., they didnt maximize the reward rate per patch; ii) they were sensitive to post-reward delays in both foraging and intertemporal choice contexts, indicating that they consider future rewards. Although rats did not behave optimally in either paradigm, their observed time preferences were more consistent with temporal discounting than ignoring the opportunity costs for these foraging and delay discounting decisions (i.e., the post-reward delay).

The idea that animals exhibit similar decision biases in foraging and intertemporal choice paradigms, and that these biases can be explained by a common model of discounting, is in conflict with prior studies that found that animals are better at maximizing long-term reward rate in foraging than in intertemporal choice tasks, and that delay discounting models of intertemporal choice tasks are poor predictors of foraging behavior (7, 16, 25, 30). It has been argued that animals may perform better in foraging tasks because decision-making systems have evolved to solve foraging problems rather than two-alternative intertemporal choice problems (15, 16, 25). However, results from the present study suggest that animals use similar rate maximizing strategies across paradigms. One potential explanation for why delay discounting models fail to predict foraging behavior is that, typically, delay discounting models only consider the decision at hand (local reward maximization). For most delay discounting tasks, this assumption is appropriate as each trial is independent and therefore, the reward value of all future decisions does not depend on the current decision. However, in patch foraging tasks, decisions are not independent — a decision to stay in the patch right now affects the potential reward on the next decision. Optimal foraging models, such as MVT, assume that animals consider future rewards, and behavior of our rats confirms this prediction. Accordingly, the temporal discounting models tested in the present study considered the discounted value of all future rewards. Furthermore, we tested a more flexible form of temporal discounting, quasi-hyperbolic discounting. This form of discounting was capable of explaining the rich pattern of behaviors observed across foraging and intertemporal choice tasks.

Although quasi-hyperbolic discounting provided the best singular explanation for rat behavior across tasks, many of the models tested were capable of explaining some of the biases exhibited by rats. We cannot exclude the possibility of subjective costs, diminishing marginal utility, or biased estimation of time intervals from contributing to suboptimal decision-making. Importantly, our data indicate that maximization of discounted future rewards may provide a link between foraging and two-alternative choice paradigms, and it highlights the importance of fu-

ture work considering the source of time preferences. These observations are buttressed by recent theoretical work demonstrating that the appearance of time preferences in standard delay discounting tasks can emerge rationally, from a value construction process whose estimates increase in variability with the delay until reward receipt (and, indeed, such as-if discounting is hyperbolic discounting when variability increases linearly with delay) (31). Further, a sequential sampling model of two-alternative forced choice (32), parameterized such that outcome delay scales variability in this way, has recently been shown to capture key dynamical features of both patch foraging (33) and hyperbolic discounting in intertemporal choice (34). Future work should build on these findings to explore directly whether the common biases identified here reflect a core computation underlying decision-making under uncertainty and across time.

Overall, the present findings elucidate decision biases exhibited by rats in foraging and intertemporal choice tasks and suggest that these biases can be explained by a common model — temporal discounting. Impairments in intertemporal choice decisions are common to a number of neuropsychiatric disorders (35). Further understanding of these decision biases is critical to understand their neural mechanisms and could help elucidate the cause of and treatment targets for neuropsychiatric disorders.

Methods

Animals

Adult Long-Evans rats were used (Charles River, Kingston, NY). One group of eight rats participated in the scale, travel time, depletion rate, and handling time experiments (in that order), a different set of eight rats were tested on the post-reward delay foraging experiment then the delay discounting task. Rats were housed on a reverse 12 h/12 h light/dark cycle. All behavioral testing was conducted during the dark period. Rats were food restricted to maintain a weight of 85-90% ad-lib feeding weight, and were given ad-lib access to water. All procedures were approved by the Princeton University and Rutgers University Institutional Animal Care and Use Committee.

Foraging Task

Animals were trained and tested as in Kane et al, 2017 (19). Rats were first trained to lever press for 10% sucrose water on an FR1 reinforcement schedule. Once exhibiting 100+ lever presses in a one hour session, rats were trained on a sudden patch depletion paradigm — the lever stopped yielding reward after 4-12 lever presses — and rats learned to nose poke to reset the lever. Next rats were tested on the full foraging task.

A diagram of the foraging task is in Fig. S1. On a series of trials, rats had to repeatedly decide to lever press to harvest reward from the patch or to nose poke to travel to a new, full patch, incurring the cost of a time delay. At the start of each trial, a cue light above the lever and inside the nose poke turned on, indicating rats could now make a decision. The time from cues turning on until rats pressed a lever or nose poked was recorded as the decision time (DT). A decision to harvest from the patch (lever press) yielded reward after a short pre-reward delay (referred to as the handling time delay, simulating the time to “handle“ prey after deciding to harvest). Reward (sucrose water) was delivered when the rat entered the reward magazine. The next trial began after an inter-trial interval (ITI). To control the reward rate within the patch, the length of the ITI was adjusted based on the DT of the current trial, such that the length of all

harvest trials was equivalent. With each consecutive harvest, the rat received a smaller volume of reward to simulate depletion from the patch. A nose poke to leave the patch caused the lever to retract for a delay period simulating the time to travel to a new patch. After the delay, the opposite lever extended, and rats could harvest from a new, replenished patch.

Details of the foraging environment for each experiment can be found in Table 1. For each experiment, rats were trained on a specific condition for 5 days, then tested for 5 days. Conditions within experiments were counterbalanced. Rat foraging behavior was assessed using mixed effects models. In the travel time experiment, we assessed the effect of starting volume of the patch and the travel time on number of harvests per patch, with random intercepts for each subject and random slopes for both variables. In all other experiments, we assessed the effect of experimental condition on harvests per patch, with random intercepts for each subject and random slopes for the effect of experimental condition.

Two-alternative choice task

Rats were immediately transferred from the foraging task to the two alternative choice task with no special training. This task consisted of a series of episodes that lasted 20 trials. At the beginning of each episode one lever was randomly selected as the shorter-sooner lever, yielding 40 µL of reward following a 1 s delay. The other lever (longer-later lever) was initialized to yield a reward of 40, 80, or 120 µL after a 1, 2, 4 or 6 s delay. For the first 10 trials of each episode, only one lever extended, and rats were forced to press that lever to learn the reward value and delay associated with the lever. The last four forced trials (trials 7-10) were counterbalanced to reduce the possibility of rats developing a perseveration bias. For the remaining 10 trials of each episode, both levers extended, and rats were free to choose the option they prefer. At the beginning of each trial, cue lights turned on above the lever indicating rats could now make a decision. Once the rat pressed the lever, the cue light turned off, and the delay period was initiated. A cue light turned on in the reward magazine at the end of the delay period, and rats received reward as soon as they entered the reward magazine. Reward magnitude was cued by light and tone. Following reward delivery, there was an ITI before the start of the next trial. At the completion of the episode, the levers retracted, and rats had to nose poke to begin the next episode (with a different larger-later reward and delay).

Two-alternative choice data was analyzed using a mixed effects logistic regression, examining the the effect of larger-later reward value and larger-later delay, both as categorical variables, on rats choices. Custom contrasts were tested using the phia package in R (36), using Holm's method to correct for multiple comparisons.

Foraging Models

All models were constructed as continuous time semi-markov processes. This provided a convenient way to capture the dynamics of timing in both tasks, including slow delivery and consumption of reward (up to 6 s for the largest rewards), and test hypotheses regarding uncertainty in time intervals. To model the foraging task, each event within the task (e.g. cues turning on/off, lever press, reward delivery, etc.) marked a state transition (abbreviated state space diagram in Fig. S2). All state transitions were deterministic, except for decisions to stay in vs. leave the patch, which occurred in 'decision' states (the time between cues turning on at the start of the trial and rats performing a lever press or nosepoke). In decision states, a decision to stay in the patch transitioned to the handling time state, then reward state, ITI state, and to the decision

state on the next trial. A decision to leave transitioned to the travel time state, then to the first decision state in the patch. Using the notation of Bradtke & Duff, 1995 (37), the reward for staying in state s, $Q(stay, s)$, is the reward provided for staying in state s, $R(stay, s)$, plus discounted value of the next state:

$$Q(stay, s) = R(stay, s) + \gamma(stay, s) * V(s_{next})$$

where $\gamma(stay, s)$ is the discount applied to the value of the next state for staying in state s, and $V(s_{next})$ is the value of the next state in the patch. For all non-decision states, rats did not have the option to leave the patch, so for these states, $V(s) = Q(stay, s)$. For decision states, the value of the state was the greater of $Q(stay, s)$ and $Q(leave)$.

For simplicity, in most models we assume the time spent in a given state is constant, calculated as the average amount of time a rat spent in the state. Under this assumption, $R(stay, s)$ is the reward rate provided over the course of the state, $r(s)$, multiplied by the time spent in that state $T(s)$, discounted according to discount factor β:

$$R(stay, s) = \frac{1 - e^{-\beta * T(s)}}{\beta} * r(s), \text{and}$$

$$\gamma(stay, s) = e^{-\beta * T(s)}.$$

The value of leaving a patch, $Q(leave)$, was equal to the discounted value of the first state in the next patch, $V(s_{first})$:

$$Q(leave) = \gamma(leave) * V(s_{first})$$

where $\gamma(leave)$ is the discount factor applied to the next state in the first patch. Assuming no variance in the travel time τ, $\gamma(leave) = e^{-\beta * \tau}$. Per MVT, we assumed rats left patches at the first state in the patch in which $Q(stay, s) \leq Q(leave)$. To capture variability in the trial at which rats left patches, we added gaussian noise to $Q(leave)$. As decisions within each patch are not independent, the patch leaving threshold did not vary trial-by-trial, but rather patch by patch, such that the cumulative probability that a rat has left the patch by state s, $\pi(leave, s)$, was the probability that $Q(stay, s) \leq Q(leave) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$, with free parameter σ.

The optimal policy for a given set of parameters was found using dynamic programming. Optimal foraging behavior is to maximize undiscounted long-term reward rate. Optimal behavior was determined by fixing the discount rate factor $\beta = .001$ and assuming no decision noise ($\epsilon = 0$). Optimal behavior was determined for each rat, in which the time spent in each state was taken from rat data. For each model, we both fit group level parameters and individual parameters for each rat using an expectation-maximization algorithm (23).

To examine linear and sublinear underestimation of post-reward delays, respectively, the time spent in post-reward delay (ITI) states was transformed, with free parameter α:

$$T_{post-linear}(s_{ITI}) = \alpha T(s_{ITI}), \text{ where } 0 < \alpha < 1, \text{or}$$

$$T_{post-sublinear}(s_{ITI}) = \frac{1 - e^{-\alpha * T(s_{ITI})}}{\alpha}, \text{ where } \alpha > 0.$$

Similarly, for overestimation of pre-reward delay, the handling time and travel time were transformed:

$$T_{pre-delay}(s_{HT}) = \alpha T(s_{HT}), \text{ and}$$

$$\tau_{pre-delay} = \alpha\tau, \text{ where } \alpha > 1.$$

For the exponential discounting model, β was fit as a free parameter.

Standard hyperbolic discounting was implemented using the µAgents model described by Kurth-Nelson & Redish, 2009 (20). The value functions of the overall model, $Q^{\mu Agent}(stay, s)$ and $Q^{\mu Agent}(leave)$, were the average of 100 µAgents, each with their own exponential discount factor β_i, and thus individual reward functions $R_i(stay, s)$, discount functions $\gamma_i(stay, s)$ and $\gamma_i(leave)$, and value functions $Q_i(stay, s)$, $Q_i(leave)$, and $V_i(s)$:

$$Q_i(stay, s) = R_i(stay, s) + \gamma_i(stay, s) * V_i(s_{next})$$

$$Q^{\mu Agent}(stay, s) = \frac{1}{100} \sum_i R_i(stay, s) + \gamma_i(stay, s) * V_i(s_{next})$$

$$Q_i(leave) = \gamma_i(leave) * V_i(s_{first})$$

$$Q^{\mu Agent}(leave) = \frac{1}{100} \sum_i \gamma_i(leave) * V_i(s_{first})$$

If the µAgent discount factors, β_i, were drawn from an exponential distribution with rate parameter $\lambda > 0$, the discounting function of the model approximated the standard hyperbolic discount function, reward$/(1 + k * \text{delay})$, with discount rate $k = 1/\lambda$. This relationship is presented in Fig. S6. λ was fit as a free parameter.

Quasi-hyperbolic discounting was originally formulated for discrete time applications (21). We used the continuous time formulation from McClure et al., 2007 (22), in which the value functions of the overall model were the weighted sum of two exponential discount systems, a steep discounting β system that prefers immediate rewards and a slower discounting δ system, each with their own reward functions, $R_\beta(stay, s)$ and $R_\delta(stay, s)$, and discount functions $\gamma_\beta(stay, s)$, $\gamma_\beta(leave)$, $\gamma_\delta(stay, s)$, and $\gamma_\delta(leave)$:

$$Q_\beta(stay, s) = R_\beta(stay, s) + \gamma_\beta(stay, s) * V_\beta(s_{next})$$

$$Q_\beta(leave) = \gamma_\beta(leave) * V_\beta(s_{first})$$

$$Q_\delta(stay, s) = R_\delta(stay, s) + \gamma_\delta(stay, s) * V_\delta(s_{next})$$

$$Q_\delta(leave) = \gamma_\delta(leave) * V_\delta(s_{first})$$

The value function of the overall quasi-hyperbolic discounting model were:

$$Q^{quasi}(stay, s) = \omega * Q_\beta(stay, s) + (1 - \omega) * Q_\delta(stay, s)$$

$$Q^{quasi}(leave) = \omega * Q_\beta(leave) + (1 - \omega) * Q_\delta(leave)$$

where $0 < \omega < 1$ was the weight of the β system relative to the δ system. We fit the parameters β, δ, and ω.

Intertemporal Choice Task Models

Similar to the foraging task, events within the task marked state transitions, and all state transitions were deterministic except for decisions to choose the smaller-sooner option (SS) or larger-later option (LL), which occurred only in decision states (abbreviated state space diagram in Fig S5A). From decision states, animals transitioned to delay, reward, and post-reward delay

(ITI) states for the chosen option — the delay, reward and ITI for the SS and LL options were represented by separate states. The value of choosing SS or LL in decision state *s* is the discounted value of the next state — the following delay state:

$$Q(SS,\ s) = \gamma(s) * Q(SS\ Delay)$$

$$Q(LL,\ s) = \gamma(s) * Q(LL\ Delay)$$

The value of delay states were the discounted value of the reward state for that action, the value of reward states were the reward for that action plus the discounted value of the ITI state for that action, and the value of ITI states were the discounted value of the decision state for the next action:

$$Q(SS\ Delay) = \gamma(SS\ Delay) * Q(SS\ Reward)$$

$$Q(SS\ Reward) = R(SS\ Reward) * \gamma(SS\ Reward) * Q(SS\ ITI)$$

$$Q(SS\ ITI) = \gamma(SS\ ITI) * V(s_{next\ dec})$$

where the value of the next decision state, $V(s_{next\ dec})$ is the greater of the value of choosing SS or choosing LL in that decision state. Decisions were made using a softmax rule, with the probability of choosing the LL option in decision state *s* defined as:

$$p(choose\ LL, s) = \frac{1}{1 + e^{\frac{Q(SS,\ s) - Q(LL,\ s)}{\sigma}}}$$

with temperature parameter σ, a free parameter that determines decision noise. The underestimating post-reward delays and temporal discounting models were implemented as they were in the foraging task.

Model Comparison

All models had two parameters except for the quasi-hyperbolic discounting model, with four. To determine the model that provided the best fit to the data, while accounting for the increased flexibility of the quasi-hyperbolic discounting model, we calculated the Bayes Information Criterion over the group level parameters (iBIC) (23, 38). iBIC penalizes the log marginal likelihood, $logp(D \mid \theta)$, which is the log probability of all the data, D, given group level parameters, θ, for model complexity. Complexity is determined by the number of parameters k, and the size of the penalty depends on the total number of observations, n:

$$iBIC = logp(D \mid \theta) + \frac{k}{2} log(n).$$

As in Huys et al., 2011, we use a Laplace approximation to the log marginal likelihood (23):

$$logp(D \mid \theta) = -\frac{n}{2} log(2\pi) * s + \sum_{i=1}^{s} p(D_i \mid \theta_i) p(\theta_i \mid \theta) - \frac{\sum_{i=1}^{s} logdet(Hf(\theta_i))}{2}$$

where s is the number of subjects, and $Hf(\theta_i)$ is the hessian matrix of the likelihood for subject i at the individual parameters θ_i.

Acknowledgements

This work was supported by NIH grant F31MH109286 (GAK) and the Princeton Program in Cognitive Science.

References

1. Charnov EL (1976) Optimal foraging, the marginal value theorem. *Theoretical population biology* 9:129–136.
2. Constantino SM, Daw ND (2015) Learning the opportunity cost of time in a patch-foraging task. *Cognitive, affective & behavioral neuroscience* 15(4):837–53.
3. Hayden BY, Pearson JM, Platt ML (2011) Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience* 14(7):933–939.
4. Nonacs P (2001) State dependent behavior and the Marginal Value Theorem. *Behavioral Ecology* 12(1):71–83.
5. Stephens DW, Krebs JR (1986) *Foraging Theory*. Vol. 121.
6. Wikenheiser AM, Stephens DW, Redish aD (2013) Subjective costs drive overly patient foraging strategies in rats on an intertemporal foraging task. *Proceedings of the National Academy of Sciences* 110(20):8308–13.
7. Carter EC, Redish AD (2016) Rats value time differently on equivalent foraging and delay-discounting tasks. *Journal of Experimental Psychology: General* 145(9):1093–1101.
8. Ainslie G (1992) *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*. (Cambridge University Press).
9. Kirby KN (1997) Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General* 126(1):54.
10. Kacelnik A (1997) Normative and descriptive models of decision making: time discounting and risk sensitivity in *Characterizing Human Psychological Adaptations*. Vol. 208, pp. 51–66.
11. Gallistel CR, Gibbon J (2000) Time, rate, and conditioning. *Psychological review* 107(2):289–344.
12. Bateson M, Kacelnik a (1996) Rate currencies and the foraging starling: The fallacy of the averages revisited. *Behavioral Ecology* 7(3):341–352.
13. Stephens DW, Anderson D (2001) The adaptive value of preference for immediacy: when shortsighted rules have farsighted consequences. *Behavioral Ecology* 12(3):330–339.
14. Stephens DW (2002) Discrimination, discounting and impulsivity: a role for an informational constraint. *Philosophical Transactions of the Royal Society B: Biological Sciences* 357(1427):1527–1537.
15. Stephens DW, Kerr B, Fernandez-Juricic E (2004) Impulsiveness without discounting: the ecological rationality hypothesis. *Proceedings of the Royal Society B: Biological Sciences* 271(1556):2459–2465.
16. Stephens DW (2008) Decision ecology: foraging and the ecology of animal decision making. *Cognitive, Affective & Behavioral Neuroscience* 8(4):475–484.
17. Pearson JM, Hayden BY, Platt ML (2010) Explicit Information Reduces Discounting Behavior in Monkeys. *Frontiers in Psychology* 1.
18. Blanchard TC, Pearson JM, Hayden BY (2013) Postreward delays and systematic biases in measures of animal temporal discounting. *Proceedings of the National Academy of Sciences* 110(38):15491–6.
19. Kane GA, et al. (2017) Increased locus coeruleus tonic activity causes disengagement from a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience*.
20. Kurth-Nelson Z, Redish AD (2009) Temporal-difference reinforcement learning with distributed representations. *PLoS One* 4(10):e7362.
21. Laibson D (1997) Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Eco-*

nomics 112(2):443–478.
22. McClure SM, Ericson KM, Laibson DI, Loewenstein G, Cohen JD (2007) Time Discounting for Primary Rewards. *Journal of Neuroscience* 27(21):5796–5804.
23. Huys QJM, et al. (2011) Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. *PLoS Computational Biology* 7(4):e1002028.
24. Huys QJM, et al. (2012) Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. *PLOS Computational Biology* 8(3):e1002410.
25. Blanchard TC, Hayden BY (2015) Monkeys Are More Patient in a Foraging Task than in a Standard Intertemporal Choice Task. *PLOS ONE* 10(2):e0117057.
26. Mazur JE (1991) Choice with Probabilistic Reinforcement: Effects of Delay and Conditioned Reinforcers. *Journal of the Experimental Analysis of Behavior* 55(1):63–77.
27. Templeton AR, Lawlor LR (1981) The Fallacy of the Averages in Ecological Optimization Theory. *The American Naturalist* 117(3):390–393.
28. Turelli M, Gillespie JH, Schoener TW (1982) The Fallacy of the Fallacy of the Averages in Ecological Optimization Theory. *The American Naturalist* 119(6):879–884.
29. Real LA (1991) Animal choice behavior and the evolution of cognitive architecture. *Science (New York, N.Y.)* 253(5023):980–986.
30. Carter EC, Pedersen EJ, McCullough ME (2015) Reassessing intertemporal choice: human decision-making is more optimal in a foraging task than in a self-control task. *Frontiers in Psychology* 6.
31. Gabaix X, Laibson D (2017) Myopia and Discounting, (National Bureau of Economic Research, Cambridge, MA), Technical Report w23254.
32. Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review* 113(4):700–765.
33. Davidson J, El-Hady A (2018) Foraging as an evidence accumulation process. *bioRxiv* p. 416602.
34. Hunter LE, Bornstein AM, Hartley CA (2018) A common deliberative process underlies model-based planning and patient intertemporal choice. *Submitted*.
35. Lempert KM, Steinglass JE, Pinto A, Kable JW, Simpson HB (2018) Can delay discounting deliver on the promise of RDoC? *Psychological Medicine* pp. 1–10.
36. de Rosario-Martinez H (2015) phia: Post-Hoc Interaction Analysis.
37. Bradtke SJ, Duff MO (1995) Reinforcement learning methods for continuous-time Markov decision problems in *Advances in neural information processing systems*. pp. 393–400.
38. MacKay DJ (2003) *Information theory, inference and learning algorithms*. (Cambridge university press).

Supporting Information

Subjective Costs Model

Similar to previous studies (6, 7), to model subjective costs, we assumed rats had an aversion to leaving a patch, possibly due to the cost of increased cognitive effort required to make a decision to leave. We formalized subjective costs using a free parameter, c, that reduced the value of the patch leaving threshold:

$$Q_{cost}(leave) = -c + \gamma(leave) * V_{cost}(s_{first}).$$

This model explained overharvesting in the travel time and depletion rate experiments, but it failed to predict later patch leaving when rats were given larger rewards with longer delays, and it failed to predict time preferences in the pre- vs. post-reward delay experiment (Fig. S3A).

Diminishing Marginal Utility Model

To investigate whether diminishing marginal utility could explain rats overharvesting behavior, we tested models in which the utility of a reward received in the task increased in a sublinear fashion with respect to the magnitude of the reward. Two different utility functions were tested: a power law function and a steeper isoelastic utility function that became increasingly risk averse with larger rewards, both with free parameter η:

$$Q_{utility}(stay, s) = U(stay, s) + \gamma(stay, s) * V_{utility}(s_{next})$$

$$U_{power}(stay, s) = R(stay, s)^{\eta}, \text{ or}$$

$$U_{isoelastic}(stay, s) = \frac{R(stay, s)^{1-\eta} - 1}{1 - \eta}.$$

Both utility models captured overharvesting in the travel time, depletion rate, and post-reward delay experiments. Furthermore, they predicted later patch leaving due to larger rewards with longer delays in the scale experiment, as larger rewards were not viewed as proportionally more valuable, but they failed to capture the magnitude of this effect. As the utility models were insensitive to the placement of delays, neither model predicted earlier patch leaving due to the pre-reward delay (Fig S3A).

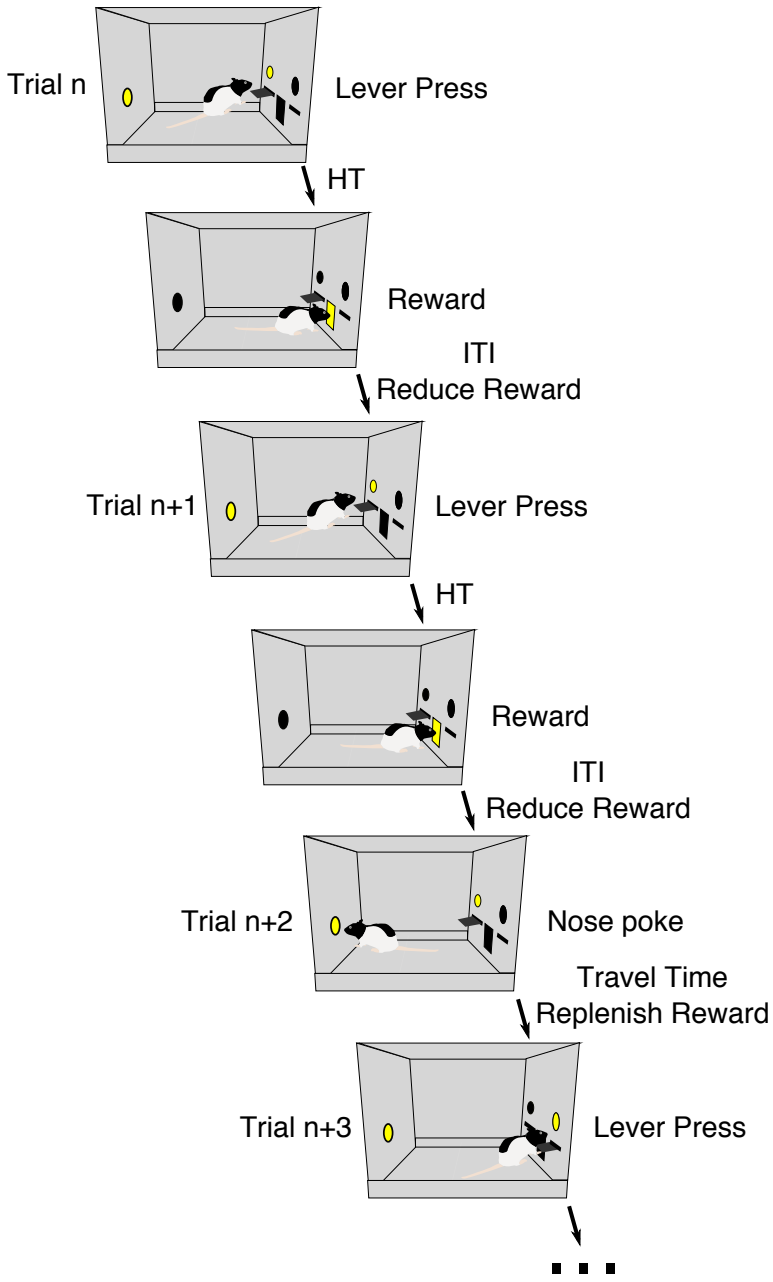

Figure S1. Diagram of the foraging task. Rats press a lever to harvest reward from the patch then receive reward in an adjacent port following a handling time delay. After receiving reward, there is an inter-trial interval (post-reward delay) before rats can make their next decision. Rats can leave the patch by nose poking in the back of the chamber (trial n+2), which initiates a delay simulating time to travel to the next patch, after which, rats can harvest from a new replenished patch.

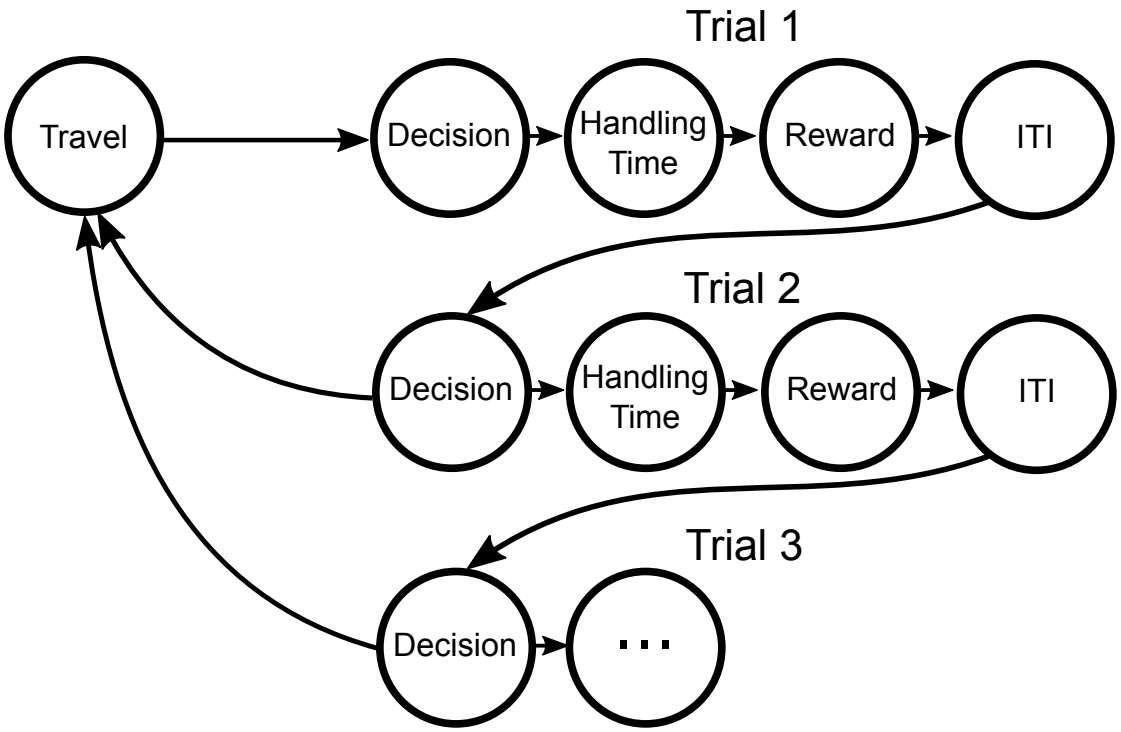

Figure S2. State space diagram for the semi-markov model of the foraging task. Decisions to stay vs. leave are made in Decision states. A Decision to stay causes a transition to the handling time, then reward, ITI, and to the Decision state on the next trial. Reward is delivered uniformly throughout time spent in the each reward state. Reward depletion is achieved via shorter time spent in reward state (resulting in longer stay in the ITI state). A Decision to leave causes a transition to the travel state, then to the first trial of the patch.

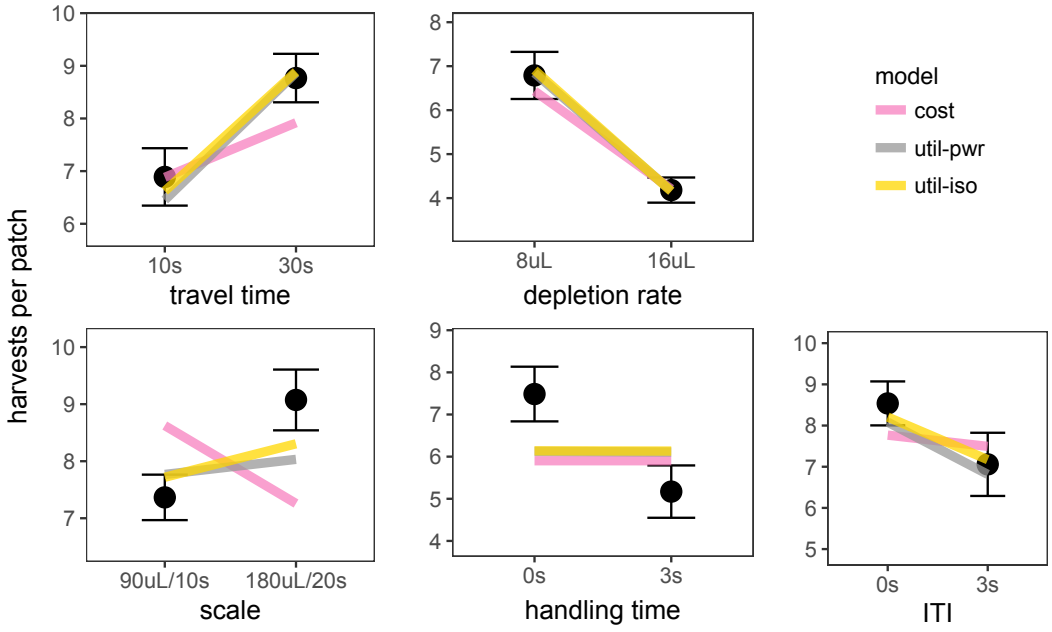

Figure S3. Predictions of the best fit subjective cost and diminishing marginal utility models (power law = util-pwr; isoelastic utility = util-iso).

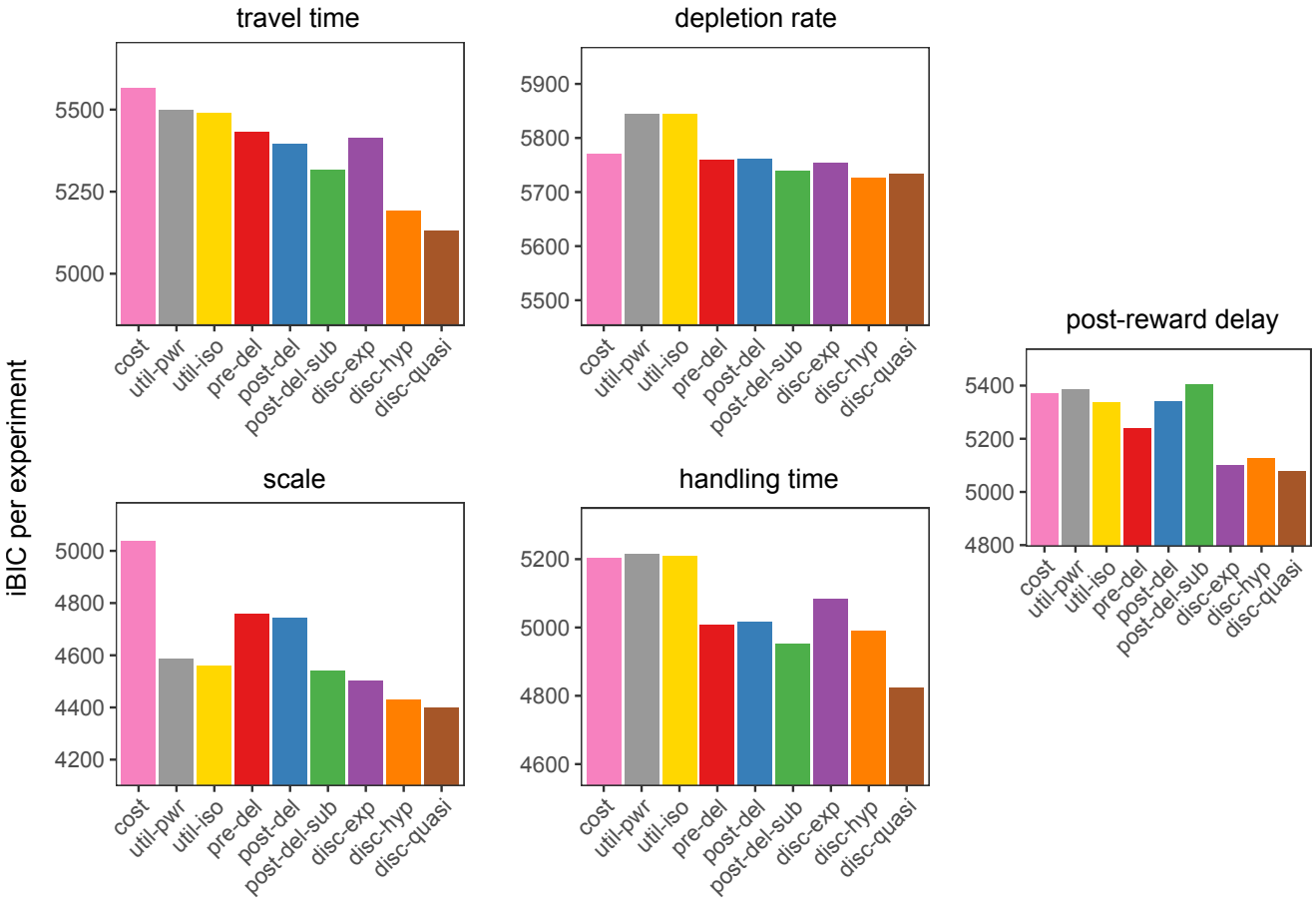

Figure S4. iBIC for each model for each foraging experiment.

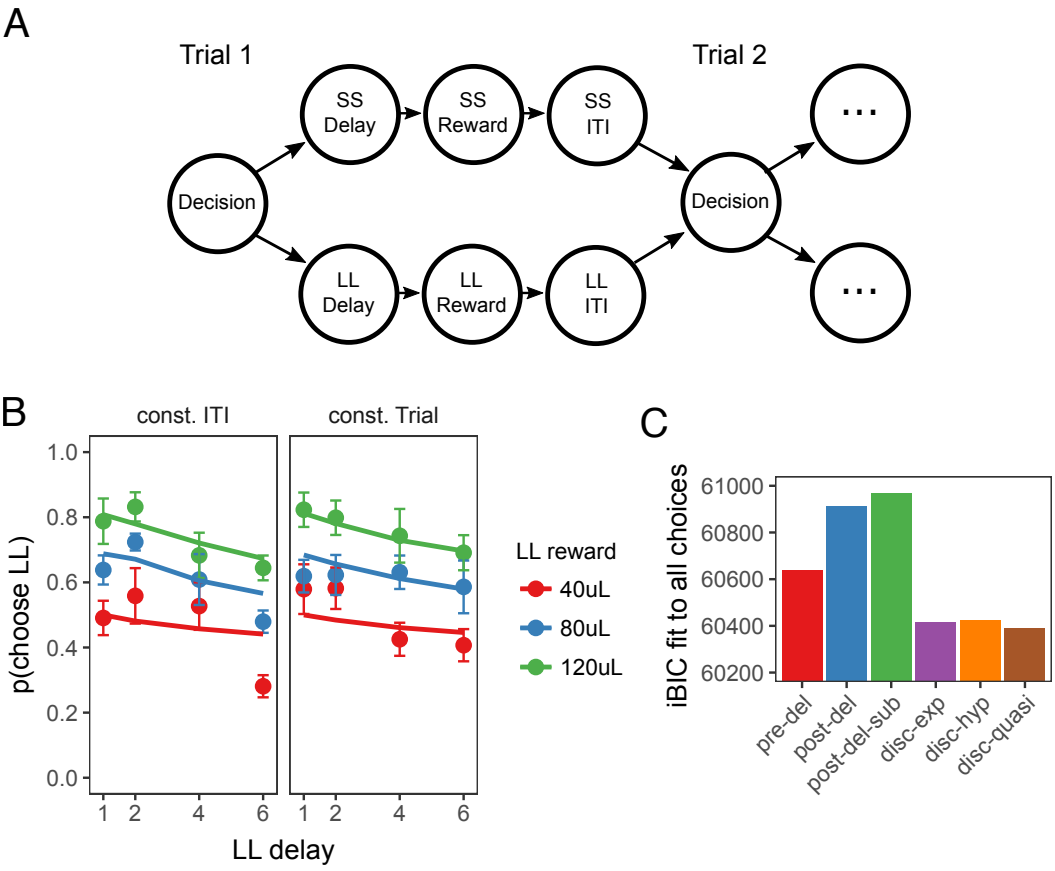

Figure S5. A) Abbreviated state space diagram for the semi-markov model of the two-alternative choice task. B) Predictions of the quasi-hyperbolic discounting model, fit only to the last five choices (as behavioral data is presented in Fig 2). Points and error bars represent mean $\pm$ standard error of rat behavior, lines represent mean model prediction. C) iBIC for each model fit to all 10 free choices.

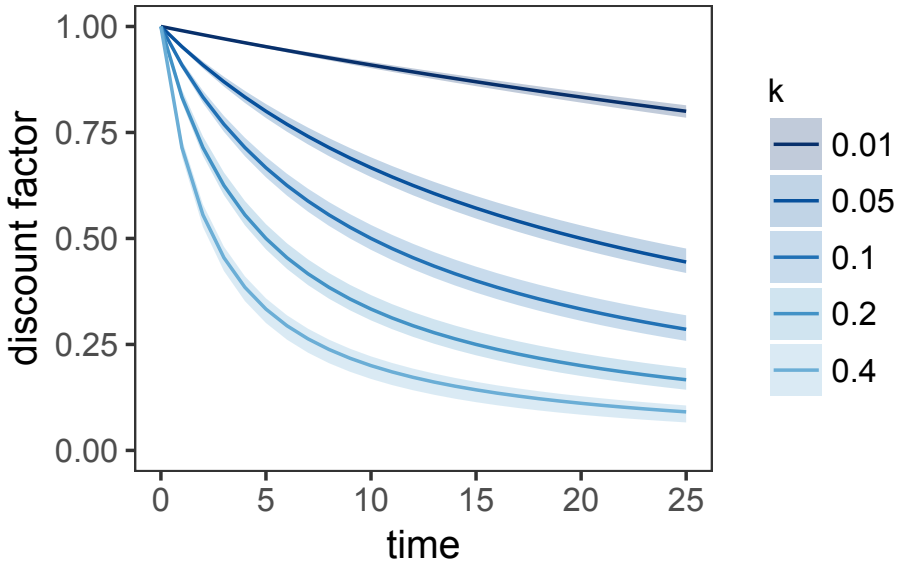

Figure S6. Discount function of the µAgent hyperbolic discounting model vs. standard hyperbolic discounting. Lines represent standard hyperbolic discounting function, $1(1 + k * time)$. Ribbon represents the mean $\pm$ standard deviation of 100 simulations of the µAgent model in which the discount factor for each of the 100 µAgents was sampled from an exponential distribution with rate parameter $\lambda = 1/k$.