Subjective costs drive overly patient foraging strategies in rats on an intertemporal foraging task

Andrew M. Wikenheiser[a], David W. Stephens[b], and A. David Redish[c,1]

[a]Graduate Program in Neuroscience and [c]Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455; and [b]Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55108

Edited by Randy Gallistel, Rutgers University, Piscataway, NJ, and approved March 29, 2013 (received for review November 28, 2012)

Laboratory studies of decision making often take the form of two-alternative, forced-choice paradigms. In natural settings, however, many decision problems arise as stay/go choices. We designed a foraging task to test intertemporal decision making in rats via stay/go decisions. Subjects did not follow the rate-maximizing strategy of choosing only food items associated with short delays. Instead, rats were often willing to wait for surprisingly long periods, and consequently earned a lower rate of food intake than they might have by ignoring long-delay options. We tested whether foraging theory or delay discounting models predicted the behavior we observed but found that these models could not account for the strategies subjects selected. Subjects' behavior was well accounted for by a model that incorporated a cost for rejecting potential food items. Interestingly, subjects' cost sensitivity was proportional to environmental richness. These findings are at odds with traditional normative accounts of decision making but are consistent with retrospective considerations having a deleterious influence on decisions (as in the "sunk-cost" effect). More broadly, these findings highlight the utility of complementing existing assays of decision making with tasks that mimic more natural decision topologies.

Intertemporal decision making treats choices between delayed outcomes and has been a topic of interest to economists, psychologists, and neuroscientists for decades (1). In certain situations, both humans and other animals commit to options that will only be realized long in the future, whereas in other settings, delay seems to erode value much more quickly (2–5). A large body of work has demonstrated that, when choosing between reinforcers of approximately equal magnitude, decision makers often prefer immediate to delayed outcomes (6–9). However, the fundamental question of how subjects compute the trade-off between the delay and magnitude of a given option remains poorly understood.

The delay discounting framework has long been invoked to both characterize and interpret intertemporal choice behavior (1). Discounting functions relate the decline of a reinforcer's subjective value with the length of time preceding delivery of that reinforcer. Initially proposed as a normative account of decision making between temporally remote outcomes, early discounting models asserted a constant decrease in subjective value with respect to delay (exponential discounting), ensuring temporally consistent decisions (10). It is now well established that discounting functions are often better described by hyperbolic curves (2, 11, 12) and that the normatively correct discounting function depends on the precise nature of the decision at hand (e.g., one-shot vs. repeated choices, probabilistic outcomes, etc.). Thus, although neither exponential nor hyperbolic discounting is universally optimal, one or the other of these models has been consistently able to fit data from myriad intertemporal choice experiments.

Intertemporal choice tasks most commonly take the form of forced choices between concurrently available options (e.g., refs. 13 and 14), a scenario infrequently encountered in nature (15). Foraging behavior is an alternative preparation for investigating intertemporal choice in circumstances akin to those that animals might encounter in natural settings (16). Searching for food is a naturally motivated behavior that has likely been subject to strong selective pressure through evolutionary time. Behavioral ecologists have formalized foraging decisions in rigorous mathematical models that provide a quantitative framework for assessing the behavioral strategies that animals adopt (17–19). Like intertemporal decision making, efficient foraging involves striking a balance between temporal costs and outcome value (8, 16, 19–21). The attractive features of the foraging framework afford great potential for integrating the behavioral ecological perspective with neuroscientific and psychological investigations of behavior, an approach that has proven fruitful previously (7, 22–26).

We designed a behavioral task to test intertemporal choice in rats. The task was constructed to mimic the natural topology of foraging choices that animals face in the wild. Because it is rare for animals to simultaneously encounter multiple potential food sources (15), foraging for food is best described as a series of go/no-go or stay/leave choices between foreground and background options (16, 19). Upon encountering a potential food source, animals decide whether to pursue and exploit that item (the foreground option), or ignore it and continue searching for other possibilities (the background). The foraging-choice topology offers a naturalistic complement to existing investigations of intertemporal choice, and provides a means of testing the validity of delay discounting models in multialternative, sequential-choice scenarios designed to approximate decision making in nature.

Results

In daily behavioral sessions, rats ($n = 10$) foraged for food pellets on a circular path outfitted with three, equidistant food pellet dispensers. Each feeder site was associated with a delay (long, medium, or short) that remained fixed at that feeder site within a session. Six combinations of delays were tested, with lengths ranging from relatively short to relatively long. These sets of delays defined six unique session types that varied in opportunity cost, or environmental richness (Fig. 1). Upon approaching a feeder location, subjects made a stay/go decision; if they remained at the site until the delay period expired, food pellets were dispensed. Otherwise, they were free to proceed to the next site at any time. Following a decision to either stay or go, that site became inactive until the rat returned to it on the subsequent lap around the track. Each feeder delivered two food pellets. Preferences were measured by computing the fraction of encounters with each feeder site in which rats waited for the delay period to expire and received food [probability of waiting (p_{wait})].

We determined the strategy that maximized food intake rate for each session type using the foraging theory prey selection model. Following the development by Stephens and Krebs (19) of Charnov's (18, 27) formulation, each feeder location was modeled as a unique prey type, with the location's delay as

Author contributions: A.M.W., D.W.S., and A.D.R. designed research; A.M.W. and A.D.R. performed research; A.M.W., D.W.S., and A.D.R. contributed new reagents/analytic tools; A.M.W. and A.D.R. analyzed data; and A.M.W., D.W.S., and A.D.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: redish@umn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1220738110/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1220738110

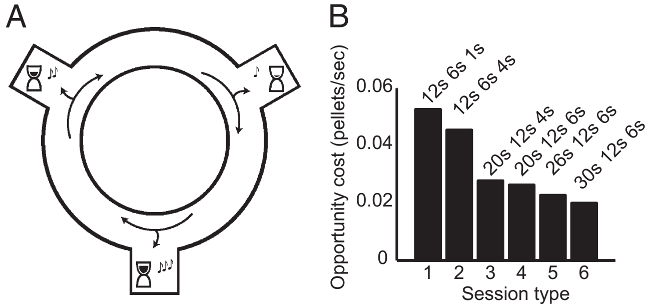

Fig. 1. (*A*) Rats foraged for food on a circular track equipped with three, equally-spaced food pellet dispenser sites. Tones cued subjects to the delay associated with each feeder location. (*B*) Task session types differed in the lengths of delays. The delays determined the session's environmental richness, or opportunity cost. When all delays were relatively short, opportunity cost was high because the environment was resource rich. As delays increased, opportunity cost fell, simulating a leaner environment.

a proxy for handling time. For our task, we found that the rate-maximizing behavior was to always accept the short-delay option, and always skip the long- and medium-length delays. Subjects did not use this strategy. Although p_{wait} generally decreased as a function of delay length (Fig. S1*A*), subjects frequently accepted both the short- and the medium-length delay options and in many cases were willing to wait for the longest delay (Fig. 2*A* and Fig. S1*B*).

To quantify performance on the task, we computed the fraction of the maximal rate rats earned within each session (the achieved rate; i.e., how well rats performed on the task compared with how well they could have performed by adopting the rate-maximizing strategy). Once again following the intuition of the diet selection model (19), we calculated the rate obtained for a given strategy as follows:

$$R = \frac{\sum_{i=1}^{3} p_{\mathrm{wait}_i}}{1 + \sum_{i=1}^{3} p_{\mathrm{wait}_i} \mathrm{delay}_i}. \qquad [1]$$

The disparity between achieved and maximal rates was striking (Fig. 2*B*). In some sessions, rats earned only 20% as much food as they might have by skipping long- and medium-length delays. At best, subjects achieved a food intake rate 25% lower than the maximum rate. These behavioral strategies clearly deviated from rate maximization.

We considered whether existing models of decision making could account for the observed behavior. The matching law (28, 29) predicts that animals adjust their level of behavioral investment to match the fraction of their total earnings (income) that each food source provides. We computed subjects' fractional investment in each option and plotted it against the fractional income subjects earned from that option; matching predicts points should be distributed along the unity line. Most observations did not conform to this prediction (Fig. 3). Although some points fell close to the matching prediction (particularly those of the medium delay option), the majority of data lay far from the diagonal. We also tested whether other aspects of behavior (dwell time distributions, the relationship between instantaneous leaving rates and overall income) were characteristic of matching strategies but found that they were not (Fig. S2), suggesting subjects were not matching.

We tested whether delay discounting could account for the behavior we observed. According to exponential discounting, the subjective value of a reinforcer falls exponentially with increasing delay (10), whereas hyperbolic discounting asserts a more concave relationship between these variables (1). We implemented Q-learning reinforcement learning (RL) models (30) that performed exponential (31) and hyperbolic (32) discounting, and tested these models on a simulation of our foraging task. Through trial and error, the simulated agents accumulated estimates of the value associated with the available actions in a given situation (the quality, or Q value of the state–action pair). The agent decided whether to wait for food delivery by comparing the Q value associated with staying at the current feeder with the Q value of proceeding to the next site. We chose the RL approach because simpler, static models require assumptions about the potentially infinite ways in which subjects might have compared options. RL models offer a simple, computationally tractable means of testing delay discounting on our task.

Behavior in the RL models depended largely on two parameters: the discounting rate (γ) and the action selection inverse temperature (β). The γ parameter controlled the rate with which value fell off as a function of delay, whereas the β parameter dictated the agent's value sensitivity (30). Small β values resulted in a strongly value-sensitive agent that tended to exploit its knowledge of the environment by strictly choosing actions with the largest Q value. Larger β values favored exploration, occasionally selecting actions with lower Q values. We tested the model over a wide range of γ and β parameters, with the aim of determining whether any combination of β and γ values could predict the behavior on the task (Fig. S3).

We computed the mean squared error (MSE) between observed and model-predicted p_{wait} values for each behavioral session (Fig. 4). Error levels were generally high, indicating a poor match between model-predicted and actual behavior. Moreover, the lowest MSE values occurred in extreme regions of the parameter space,

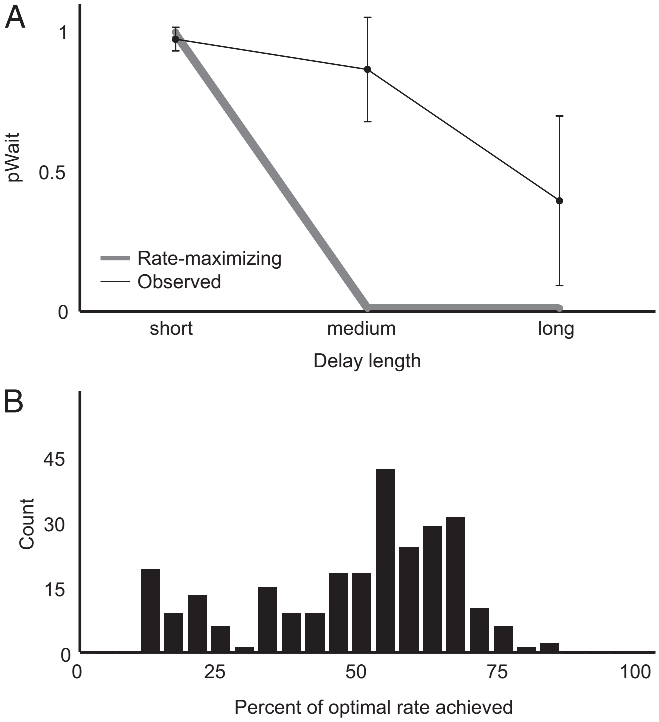

Fig. 2. (*A*) Observed p_{wait} are plotted in black, and the rate-maximizing p_{wait} are shown in gray. Error bars indicate the SD (n = 10 rats, 4 repetitions per subject of each session type). Subjects often waited for the medium and long delays, although the rate-maximizing strategy in each case was to accept only the short-delay option. Data here are collapsed across session type. Plots for each session type are shown in Fig. S1. (*B*) We computed the fraction of the maximal rate of food intake subjects earned on the task (n = 240 sessions). The achieved rate was generally substantially lower than the rate-maximizing strategy. At best, subjects earned nearly 25% less food than they would have with the rate-maximizing solution.

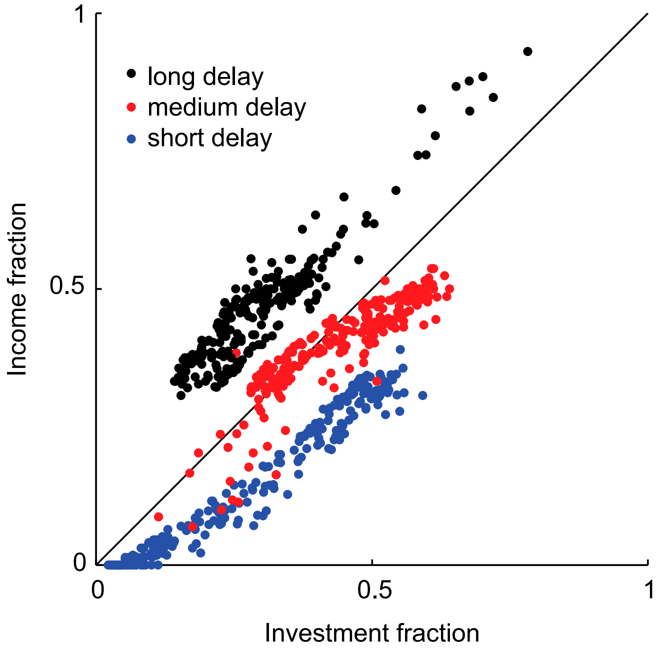

Fig. 3. Matching law predicts that the behavioral investment fraction for a given option should equal (match) the fraction of income earned from that option. Accordingly, when income and investment fractions are plotted against each other, matching predicts points should fall along the black unity line. However, the behavior we observed deviates substantially from this prediction.

corresponding to agents that seem unrealistic from a behavioral perspective; models with such large γ values show little sensitivity to delay, in contrast with rats' behavior on the task (Fig. 2 and Fig. S1). Together, the hyperbolic and exponential RL models suggest that temporal discounting cannot explain rats' behavior on the task.

The behavioral data suggest that rats were not maximizing their rate of food intake as prescribed by the foraging theory prey model (Eq. **1**); nor were rats using a matching strategy, or discounting delayed reward in accord with economic theory. To better understand how subjects arrived at the strategies they selected, we developed a version of the foraging theory prey selection model (19) that predicted the decisions we observed. Rats' propensity to wait for long- and medium-length delays was the fundamental discrepancy between observed behavior and the prey model predictions; rats behaved as though there were some cost to rejecting prey types. To capture this notion quantitatively, we modified Eq. **1** to include an aversion parameter A, representing an unwillingness to reject potential food options upon encounter:

$$R_s = \frac{\sum_{i=1}^{3} p_{\text{wait}_i} - \sum_{i=1}^{3}\left(1 - p_{\text{wait}_i}\right)A}{1 + \sum_{i=1}^{3} p_{\text{wait}_i}\,\text{delay}_i}. \qquad \textbf{[2]}$$

Here, the subjective rate of food intake (R_s) decreases with instances of prey rejection, in proportion to the A parameter. Larger A values result in the model exhibiting greater aversion to rejecting prey types. Hereafter, we refer to Eq. **2** as the "rejection-averse" rate equation, to contrast it with the standard rate equation (Eq. **1**).

Including the A parameter strongly impacted the task's reward structure. We used the rejection-averse rate equation to calculate R_s for all possible strategies in all session types. Fig. 5 shows how R_s varies across strategies, with increasing values of A. When A is zero, the model reduces to the standard rate equation prediction, with high-rate strategies concentrated in the behavioral space of accepting only the short-delay option. However, as A increases, a larger volume of strategy space yields relatively high rates of food intake.

Given the large changes in task reward structure brought about by the rejection-averse rate equation, when $A > 0$, behavioral strategies should differ substantially from the predictions of the standard rate model. Consistent with this idea, when $A = 0$, subjects' behavior (marked with black dots, Fig. 5) falls well outside of the high-rate region, but as A increases, the profitable region of strategy space shifts to encompass the behavior rats showed on the task. This suggests that subjects may have been behaving in accordance with the rejection-averse equation. To test this idea, we assumed that subjects chose strategies according to Eq. **2** and found the value of A that maximized the achieved rate of food intake for each session. Rate-maximizing A values were greater than zero ($P_{A=0} = 4.25 \times 10^{-42}$, sign rank test).

We compared the achieved rates of food intake (the fraction of the maximal rate) subjects earned in each session, assuming their behavior was guided by either the rejection-averse or the standard rate equation (Fig. 6, "observed behavior"). Behavior governed by the rejection-averse model earned a substantial fraction of the maximum possible rate (achieved rate, 95 ± 5%). As shown previously, however (Fig. 2*B*), behavior in terms of the unmodified rate equation was quite poor (achieved rate, 47 ± 18%). This suggests that subjects might have computed their rate of food intake in light of a subjective aversion to abandoning potential food sources (Eq. **2**) rather than the rate-maximizing, cost-free perspective (Eq. **1**).

Demonstrating that subjects' behavior nearly maximizes R_s (with the appropriate A parameter) shows that there exists a subjective valuation system (i.e., Eq. **2**) that can account for rats' behavior on the task. However, the rejection-averse equation contains one more parameter that the unmodified rate equation; thus, when comparing the two models it is important to consider whether any arbitrary behavioral strategy would likewise appear optimal, given the correct A value (33). We fit A for the behavior predicted by the exponential and hyperbolic discounting RL models that most closely matched subjects' actual behavior. Achieved rates are plotted in Fig. 6 with respect to the rejection-averse and standard prey selection frameworks. For both the exponential and hyperbolic models, including the A parameter increased the achieved rates associated with model behavior (from 32 ± 11% to 52 ± 7% for the exponential model; from 55 ± 16% to 72 ± 8% for the hyperbolic model). However, in neither case did the achieved rates approach those of observed behavior under the rejection-averse assumption ($P_{\text{Robserved=Rexponential}} = 1.00 \times 10^{-44}$; $P_{\text{Robserved=Rhyperbolic}} = 1.46 \times 10^{-44}$; rank sum tests).

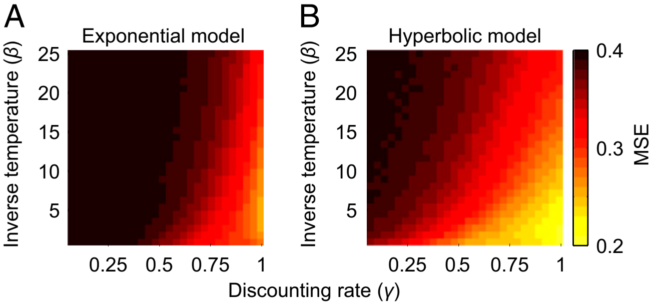

Fig. 4. To test whether any RL model parameter combinations produced behavior similar to that of subjects, we computed the mean squared error (MSE) between model predictions and observed behavior for all parameter combinations and averaged across sessions (n = 240 sessions). For both exponential (*A*) and hyperbolic (*B*) models, best fits were achieved with parameters that correspond to little or no temporal discounting.

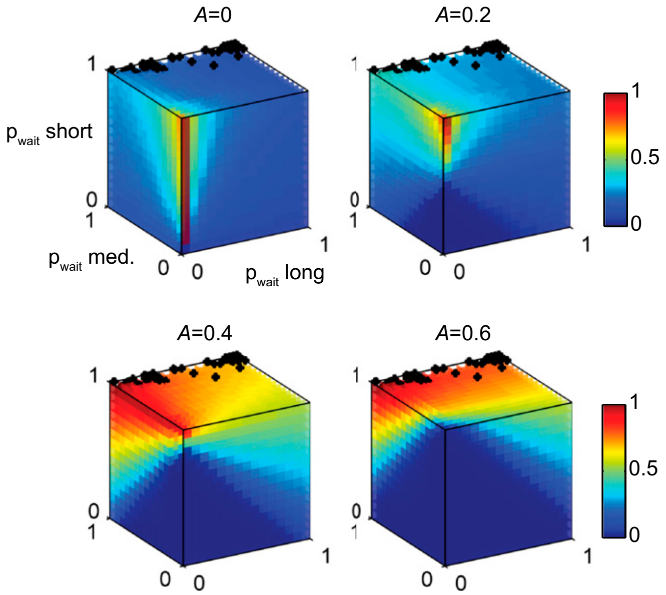

Fig. 5. We calculated the subjective rate of reward for all possible behavioral strategies using Eq. **2** and colored each region of strategy space with the corresponding normalized rate. For low values of *A*, the predictions of foraging theory hold, and high-rate strategies correspond to accepting only short delays. With increasing values of *A*, high-rate strategies shift to different regions of the strategy space. When $A = 0$, subjects' behavior (marked with black dots, $n = 40$ sessions) falls far from the strategies with high rates of food intake. However, as the region of high-rate strategies shifts with increasing values of *A*, subjects' behavior falls in much more profitable regions of strategy space.

Thus, although including the *A* parameter generally resulted in an increased achieved rate, this increase did not drive the achieved rates for all possible strategies to maximal values, suggesting that Eq. **2** is not so inappropriately flexible as to maximize intake rates for any strategy.

Interestingly, the *A* parameter fit to subjects' behavior varied across session types. The best-fitting *A* value correlated positively with the opportunity cost of time, which differed across session types (Fig. 7; $P_{\mathrm{corr}=0} = 7.77 \times 10^{-33}$, $R^2 = 0.42$). This correlation suggests that rats' sensitivity to rejection costs was not a static quantity but was instead dynamically modulated by the reward statistics of the environment. With increasing opportunity cost, rats grew more averse to abandoning potential food items. This is the reverse of the expected relationship between behavioral vigor and opportunity cost (34, 35), but is consistent with parameter *A* reflecting the opportunity lost getting to the reward (retrospective evaluation) rather than the opportunity gained or lost leaving the reward site early (prospective evaluation).

Discussion

While performing a foraging task, rats chose behavioral strategies that resulted in a large loss of food earnings compared with the maximal possible intake rates. We tested models from foraging theory, psychology, and economics, but they could not explain the behavior we observed. Importantly, delay discounting models, for many years the dominant means of evaluating and modeling intertemporal decisions, were unable to account for the preferences subjects exhibited. To model subjects' behavior, we modified the foraging theory prey selection rate equation to include a cost for rejecting potential prey items. This rejection-averse model closely matched rats' behavior (but not behavior predicted by temporal discounting models), suggesting it might capture an aspect of the decision-making process rats used on the foraging task.

What is the nature of the perceived cost that influenced rats' behavior on the foraging task? One possibility is that *A* represents some general movement or movement initiation cost. However, because the physical dimensions of the foraging task apparatus were unchanged across session types, any parameter related generally to movement costs should be constant across session types, in contrast to our findings (Fig. 7). This suggests that *A* cannot be fully explained by movement costs or other general energetic expenditures.

The *A* parameter fit to subjects' behavior varied systematically across session types and was positively correlated with the opportunity cost of the session. Opportunity cost quantifies the average value of time, given the density of reinforcement available in the environment (34). When environmental resources are abundant, opportunity cost (i.e., the cost of allocating time to an activity) is high. When resources are scarce, less reward per unit time is at stake, and opportunity cost is low. Our data conflict with the normative prediction that increasing opportunity costs ought to "invigorate" behavior (35). In sessions where the value of time was greatest (high opportunity cost), subjects were most willing to wait out long-delay options (high *A* values). Conversely, when opportunity cost was low (and subjects stood to lose little by accepting low-rate items), rats' aversion to rejecting feeder sites was lower. These findings seem inconsistent with current thinking on how opportunity cost ought to influence behavior (34, 35).

Because behavior driven by the rejection-averse rate model diverged so strikingly from rate maximization, we considered whether the *A* parameter might map on to some bias known to affect human decision makers (25). One such bias is the sunk-cost fallacy, an economic error in which willingness to continue pursuit of an option is influenced by past investment in that option, rather than anticipated future returns (36, 37). In economic terms, considering sunk costs is irrational, as investments committed to a course of action cannot be recovered. Nevertheless, in many cases decision makers show stronger preference for options they have invested resources in, despite the poor long-term consequences this might entail (37, 38).

How might the sunk-cost effect manifest on our task? Consider the decision a subject faces upon arriving at the long-delay feeder site: a forward-thinking, rate-maximizing rat would skip that feeder, proceeding to a shorter delayed site instead, and thereby enjoying a greater overall rate of food earnings. In contrast,

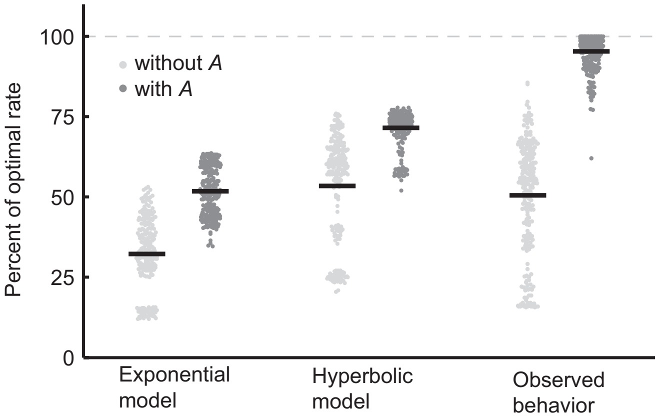

Fig. 6. We computed the achieved rates of food intake (the fraction of the maximal possible rate) for observed behavior and the best-fitting RL models, using both the rejection-averse rate equation (Eq. **2**) and the standard rate model (Eq. **1**). In general, the rejection-averse model outperformed the standard rate model in terms of achieved rates. However, the RL models' rates did not approach actual behavior with the rejection-averse assumption, suggesting Eq. **2** does not maximize R_s for any arbitrary behavioral strategy.

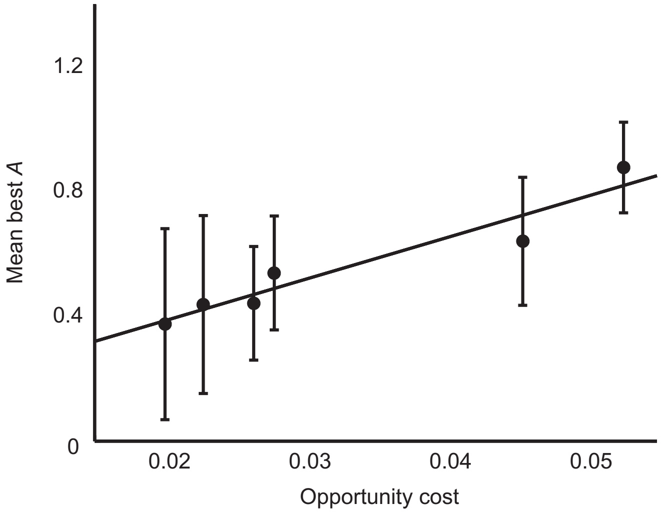

Fig. 7. We found the best-fitting *A* parameter for each behavioral session (n = 240 sessions), and measured how the mean *A* parameter varied across session types. *A* was correlated with the opportunity cost of time, which varied across session types ($P = 7.77 \times 10^{-33}$, $R^2 = 0.42$).

a rat making its decision with respect to the cost already spent getting to its current position would be more likely to wait out the delay, despite the consequent decrease in overall food intake rate. Thus, in our task, the extent to which subjects were sensitive to sunk costs is indexed by how frequently they accepted feeder sites that decreased their overall food intake. The *A* parameter, then, could be considered a session by session measure of sunk-cost sensitivity on the foraging task, where larger *A* values indicate a stronger influence of sunk costs on decision making.

The influence of sunk costs can also explain why *A* was positively correlated with opportunity cost (contrary to normative predictions) (34, 35). In addition to energy, rats also invested time in traveling between and waiting at feeders. Although the distance between food options was fixed, perceived differences in the value of time across session types would result in rats' subjective valuation of a consistent time investment increasing as opportunity cost grew. When opportunity cost was high, the subjective behavioral investment (i.e., the sunk cost) would have seemed greater. Thus, for decision makers succumbing to the sunk-cost fallacy, reluctance to abandon potential food items would grow with increasing opportunity cost, consistent with our observations (Fig. 7).

Why are decision makers swayed by retrospective investment? An influential account (38, 39) suggests that sunk costs affect behavior because humans inappropriately overgeneralize an aversion to wasting valuable resources. Although it is generally a good strategy to avoid squandering valuable resources, misapplication of the waste-aversion heuristic could lead to continued investment in a doomed venture, because sticking with a losing option subjectively validates previous investment. Complementary theories have suggested that humans attend to sunk costs to maintain their reputation (either to themselves or others) as self-consistent decision makers that avoid wasting retrospective allocations, leading to the puzzling scenario in which a misguided attempt to appear rational leads to demonstrably suboptimal behavior (40–42). One feature all these theories share is a critical role for complex social, cognitive, or metacognitive explanatory mechanisms; accordingly, these theories explicitly predict that animals either have cognitive mechanisms similar to those responsible for the sunk-cost effect in humans, or that they should be immune to biases induced by sunk costs (38). The data presented here provide evidence that retrospective considerations might influence the behavior of rats in a manner consistent with the sunk-cost fallacy, suggesting that the waste-aversion theory of human sunk-cost sensitivity (38, 39) might have deep evolutionary roots, possibly reflecting evolutionary elaboration of a simpler, learned correlation between effort invested and return.

Other psychological factors may explain rats' behavior on the task. For instance, if subjects perceived uncertainty in food delivery, waiting out long delays might have been informative, leading to a better understanding of the task's reward structure. Other species actively seek reward-related information in a behavioral task, even when this information cannot influence their earnings (43). A similar effect in rats might account for their oversampling of long delays. However, because the task was well learned, and there was no evidence for within-session changes in strategies, it seems unlikely that behavior was strongly affected by uncertainty.

Perceived uncertainty could also influence preferences in other ways. If, for instance, rats perceived food at their current location as somehow more likely to be delivered than food at future sites, such beliefs could drive a heuristic where "a bird in the hand" is valued more highly than hypothetical future reward, despite differences in delay. Finally, it is possible that subjects were not sensitive to reward rate and were instead tracking their food earnings by some other, as-yet-unknown metric.

Interestingly, previous laboratory studies of foraging decisions have found that behavior often approximates the optimal solution to the decision problem at hand (reviewed in refs. 16 and 19). For instance, birds selecting prey items from a conveyer belt (44), blue jays making stay/go decisions between food patches (45), and primates making virtual patch residence choices (24) all matched the normative predictions of foraging theory models. Our experiment differs from many previous investigations of foraging behavior in at least one critical aspect: rats performing our task were required to physically move between spatially segregated feeder sites to indicate their decisions. Our findings suggest that, at least in some cases, virtual search costs (simulated by delays) may not be equivalent to the actual energetic expenditures of moving through an environment in pursuit of food.

Methods

Subjects. Male Fisher–Brown-Norway hybrid rats (n =10; Harlan), aged 10–14 mo, were maintained on a 12-h light–dark cycle. Behavioral sessions occurred at the same time daily, during the light phase. Rats were handled 7 d and acclimated to eating 45-mg sucrose pellets (Test Diet) before beginning training. Subjects were food deprived to no less than 80% of their free-feeding weight; water was always freely available in the home cage. All experimental and animal care procedures complied with National Institutes of Health guidelines for animal care and were approved by the Institutional Animal Care and Use Committee at the University of Minnesota.

Apparatus. Rats performed the task on an elevated, circular track (width, 10 cm; diameter, 80 cm). Food pellets (Research Diets), were delivered by automated dispensers (Med Associates). An overhead camera recorded the subjects' position via a light-emitting diode (LED) "backpack," a cloth strip containing a battery-powered LED fastened around the rat's body with Velcro. Online position tracking was processed with a Cheetah 160 data acquisition system (Neuralynx). The task was controlled by custom Matlab (MathWorks) software.

Training and Task. Subjects were first trained to run clockwise around the track for food at each feeder. Attempts to run counterclockwise were blocked by the experimenter during training sessions (during task sessions, rats ran only clockwise, so blocking was unnecessary). After rats ran 30 or more laps for three consecutive sessions, the training phase was considered complete and task performance began.

During each daily, 30-min session of the foraging task, rats could earn sucrose pellets from the three feeder locations after a delay period. The delay began when the subject entered a 15-cm zone centered on the feeder site. Entry into this zone was signaled by a tone sequence (200-ms pulses, repeated once per second). The frequency of the tone was proportional to the site's delay, to inform subjects of the wait length. For two rats, tones decreased in frequency with each pulse until food delivery. For the remaining eight rats, the frequency was constant throughout the delay. No behavioral differences were observed between the two conditions, so subjects were pooled for analysis. Excluding the animals who experienced a decreasing tone did not change the results presented here. Six combinations of delays defined six unique session types. Rats experienced session types in pseudorandom order (the same type was never repeated on consecutive days). Delays were counterbalanced across feeder sites to ensure that delay distributions at each location were equivalent across sessions.

Data Analysis and Modeling. All analyses and statistical tests were conducted using Matlab. The exponential discounting RL model was implemented after Watkins (31), and the hyperbolic discounting model was implemented after Kurth-Nelson and Redish (32) (*SI Appendix*). With experience, such models learn the value of taking actions in various world states (*Q* values for state–action pairs). Action selection in both models was achieved via a soft-max algorithm (30) that compared the value associated with staying at a feeder site (the discounted value at the current site) with the value of proceeding to the next location (the delay plus travel time-discounted value of the next site). The β parameter set agents' value sensitivity, whereas the γ parameter determined how quickly subjective value decreased with delay (30).

For matching analyses, fractional investment was the sum of time spent waiting at each site divided by the total time spent waiting at all feeder sites. Similarly, fractional income was the number of pellets earned from each feeder divided by the total number of pellets earned in the session.

ACKNOWLEDGMENTS. We thank Zeb Kurth-Nelson and members of the A.D.R. Laboratory for helpful discussions on this work. This work was supported by National Institutes of Health (NIH) Grant T32-DA-070106-05 (to A.M.W.), a University of Minnesota Doctoral Dissertation Fellowship (to A.M.W.), NIH Grant R01-MH-080318 (to A.D.R.), and National Science Foundation Grant 0727221 (to D.W.S.).

1. Madden G, Bickel W, Critchfield T, eds (2009) *Impulsivity: Theory, Science, and Neuroscience of Discounting* (APA Books, Washington, DC).
2. Ainslie G (1975) Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychol Bull* 82(4):463–496.
3. Herrnstein RJ, Prelec D (1991) Melioration: A theory of distributed choice. *J Econ Perspect* 5(3):137–156.
4. Frederick S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: A critical review. *J Econ Lit* 40(4):351–401.
5. Heyman G (2009) *Addiction: A Disorder of Choice* (Harvard Univ Press, Cambridge, MA).
6. Stephens DW, Kerr B, Fernández-Juricic E (2004) Impulsiveness without discounting: The ecological rationality hypothesis. *Proc Biol Sci* 271(1556):2459–2465.
7. Kacelnik A (2003) The evolution of patience. *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, eds Loewenstein G, Read D, Baumeister R (Russell Sage Foundation, New York), pp 115–138.
8. Stevens JR, Hallinan EV, Hauser MD (2005) The ecology and evolution of patience in two New World monkeys. *Biol Lett* 1(2):223–226.
9. Hayden BY, Platt ML (2007) Animal cognition: Great apes wait for grapes. *Curr Biol* 17(21):R922–R923.
10. Samuelson PA (1937) A note on measurement of utility. *Rev Econ Stud* 4(2):155–161.
11. Ainslie G (1992) *Picoeconomics* (Cambridge Univ Press, Cambridge, UK).
12. Redish AD, Kurth-Nelson Z (2010) Neural models of temporal discounting. *Impulsivity: The Behavioral and Neurological Science of Discounting*, eds Madden G, Bickel W (APA Books, Washington, DC), pp 123–158.
13. Mazur J (1987) An adjusting procedure for studying delayed reinforcement. *Quantitative Analyses of Behavior: The Effect of Delay and of Intervening Events*, eds Commons M, Mazur J, Nevin J, Rachlin H (Erlbaum, Hillsdale, NJ), pp 55–73.
14. Cardinal RN, Daw N, Robbins TW, Everitt BJ (2002) Local analysis of behaviour in the adjusting-delay task for assessing choice of delayed reinforcement. *Neural Netw* 15(4–6):617–634.
15. Kacelnik A, Vasconcelos M, Monteiro T, Aw J (2011) Darwin's "tug of war" vs. starlings "horse-racing": How adaptations for sequential encounters drive simultaneous choice. *Behav Ecol Sociobiol* 65(3):547–558.
16. Stephens DW (2008) Decision ecology: Foraging and the ecology of animal decision making. *Cogn Affect Behav Neurosci* 8(4):475–484.
17. MacArthur R, Pianka E (1966) On optimal use of a patchy environment. *Am Nat* 100(916):603–609.
18. Charnov EL (1976) Optimal foraging, the marginal value theorem. *Theor Popul Biol* 9(2):129–136.
19. Stephens D, Krebs J (1986) *Foraging Theory* (Princeton Univ Press, Princeton).
20. Kalenscher T, Pennartz CMA (2008) Is a bird in the hand worth two in the future? The neuroeconomics of intertemporal decision-making. *Prog Neurobiol* 84(3):284–315.
21. Stevens J (2010) Intertemporal choice. *Encyclopedia of Animal Behavior*, eds Breed M, Moore J (Academic, London), pp 203–208.
22. Kacelnik A, Todd I (1992) Psychological mechanisms and the marginal value theorem: Effect of variability in travel time on patch exploitation. *Anim Behav* 43(2):313–322.
23. Pompilio L, Kacelnik A (2010) Context-dependent utility overrides absolute memory as a determinant of choice. *Proc Natl Acad Sci USA* 107(1):508–512.
24. Hayden BY, Pearson JM, Platt ML (2011) Neuronal basis of sequential foraging decisions in a patchy environment. *Nat Neurosci* 14(7):933–939.
25. Freidin E, Kacelnik A (2011) Rational choice, context dependence, and the value of information in European starlings (*Sturnus vulgaris*). *Science* 334(6058):1000–1002.
26. Kolling N, Behrens TE, Mars RB, Rushworth MF (2012) Neural mechanisms of foraging. *Science* 336(6077):95–98.
27. Charnov E (1976) Optimal foraging: Attack strategy of a mantid. *Am Nat* 110(971):141–151.
28. Herrnstein RJ (1970) On the law of effect. *J Exp Anal Behav* 13(2):243–266.
29. Gallistel CR, Mark TA, King AP, Latham PE (2001) The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *J Exp Psychol Anim Behav Process* 27(4):354–372.
30. Sutton RS, Barto AG (1998) *Reinforcement Learning: An Introduction* (MIT, Cambridge, MA).
31. Watkins C (1989) Learning from delayed rewards. PhD thesis (Cambridge University, Cambridge, UK).
32. Kurth-Nelson Z, Redish AD (2009) Temporal-difference reinforcement learning with distributed representations. *PLoS One* 4(10):e7362.
33. Roberts S, Pashler H (2000) How persuasive is a good fit? A comment on theory testing. *Psychol Rev* 107(2):358–367.
34. Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology (Berl)* 191(3):507–520.
35. Niv Y, Joel D, Dayan P (2006) A normative perspective on motivation. *Trends Cogn Sci* 10(8):375–381.
36. Aronson E (1961) The effect of effort on the attractiveness of rewarded and unrewarded stimuli. *J Abnorm Soc Psychol* 63(2):375–380.
37. Arkes H, Blumer C (1985) The psychology of sunk cost. *Organ Behav Hum Decis Process* 35(1):124–140.
38. Arkes H, Ayton P (1999) The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychol Bull* 125(5):591–600.
39. Arkes H (1996) The psychology of waste. *J Behav Decis Making* 9(3):213–224.
40. Staw B (1976) Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organ Behav Hum Perform* 16(1):27–44.
41. Staw B, Fox F (1977) Escalation: The determinants of commitment to a chosen course of action. *Hum Relat* 30(5):431–450.
42. Brockner J (1992) The escalation of commitment to a failing course of action: Toward theoretical progress. *Acad Manage Rev* 17(1):39–61.
43. Bromberg-Martin ES, Hikosaka O (2009) Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* 63(1):119–126.
44. Krebs J, Erichsen J, Webber M (1977) Optimal prey selection in the great tit (*Parus major*). *Anim Behav* 268(5616):30–38.
45. Stephens D, Anderson D (2001) The adaptive value of preference for immediacy: When shortsighted rules have farsighted consequences. *Behav Ecol* 12(3):330–339.

Supporting Information

Wikenheiser et al. 10.1073/pnas.1220738110

SI Appendix

Testing for Matching Behavior. In addition to the analyses reported in the main text (Fig. 3), we examined several other aspects of behavior for characteristics of the matching strategy. On matching tasks, the distributions of visit durations (dwell times) for each option are typically exponentially distributed. Because different feeder sites in our task were associated with different delay lengths, we normalized visit durations by each site's delay length to view all of the data on the same scale (Fig. S2*A*). In our task, subjects sometimes waited for the delay period to expire, but in other cases left the site before this time. Accordingly, the overall distribution of visit durations had two peaks—one resulting from trials in which subjects waited for food delivery and another resulting from trials in which subjects left the site early. To fairly consider whether visit durations were distributed exponentially, we considered the complete set of visits (all trials), and also separated trials based on whether subjects waited for food delivery (wait trials only) or left before earning food (skip trials only). We computed survivor curves (1) for visit durations separated in this way and compared them with survivor curves calculated for an exponential distribution with the mean matched to the mean visit duration of the sample in question (Fig. S2*B*). Examining these plots shows substantial divergence between the empirical exponential survivor curves (dashed lines) and the visit duration survivor curves (solid lines), suggesting that visit durations were not distributed exponentially.

To ensure that normalizing and pooling data did not affect our analysis, we directly tested distributions of visit durations from each site within each session for exponentiality using a Lilliefors test of the null hypothesis that sample data derive from an exponential distribution with an unspecified mean. Again, we separately considered visit durations from all trials, skip trials only, and wait trials only. To ensure that enough samples were present to accurately assess the distribution's shape, only distributions with at least 10 or more samples were tested. We found that, in the vast majority of cases (all trials: 99%; skip trials only: 96%; wait trials only: 100%), there was sufficient evidence to reject the null hypothesis that data were distributed exponentially ($P < 0.05$). Together, these results demonstrate that visit durations did not follow an exponential distribution.

We also examined the relationship between the sum of the average leaving rates (the reciprocal of the average dwell time for each site in a session) and the sum of incomes for each site. Matching predicts a linear relationship between these quantities; however, we observed no clear relationship in our data (Fig. S3).

Taken together, these analyses suggest rats did not use a matching strategy while performing the foraging task.

Reinforcement Learning Models. We tested a wide range of β and γ values; each parameter combination was simulated 10 times and behavior (p_{wait}) was averaged across repetitions. Travel time between feeder sites was fixed at 2 s, a time consistent with rats' behavior on the task. For both models, the learning rate was fixed at a moderate value (0.15), and behavior was examined after simulating 2,000 time steps, when *Q*-value estimates had stabilized.

Exponential Discounting Reinforcement Learning Model. Delay was discounted exponentially in the exponential model, following SV = Mγd [SV = subjective value, M = reinforcer magnitude, d = delay, γ (∈ [0, 1]) = discounting rate].

Microagents Hyperbolic Discounting Reinforcement Learning Model. To model hyperbolic discounting, we used a previously described Q-learning model (2) in which behaviorally relevant variables were represented across a population of independent "microagents." Each microagent ($n = 1{,}000$) computed its own estimate of *Q* values and discounted delayed reward exponentially, according to its own, unique γ parameter. The "macroagent" averaged *Q* values over the microagent distribution when making decisions and exhibited hyperbolic discounting as an emergent property of interactions within the population of microagents (2). Action selection was achieved with the same soft-max algorithm as in the exponential model (3). To change the delay tolerance of the macroagent, we altered the distribution of microagent discounting rates by raising each value in the population to an exponent ranging from 0.01 to 100; distributions skewed toward high γ values resulted in slower discounting, whereas a preponderance of low γ values resulted in faster discounting. Consequently, the macroagent, unlike the exponential model, is not characterized by a single discounting rate; instead, in Fig. 4 and Fig. S3, axes for the hyperbolic model display the median γ value of the microagent distribution.

1. Gallistel CR, Mark TA, King AP, Latham PE (2001) The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *J Exp Psychol Anim Behav Process* 27(4):354–372.
2. Kurth-Nelson Z, Redish AD (2009) Temporal-difference reinforcement learning with distributed representations. *PLoS One* 4(10):e7362.
3. Sutton RS, Barto AG (1998) *Reinforcement Learning: An Introduction* (MIT, Cambridge, MA).

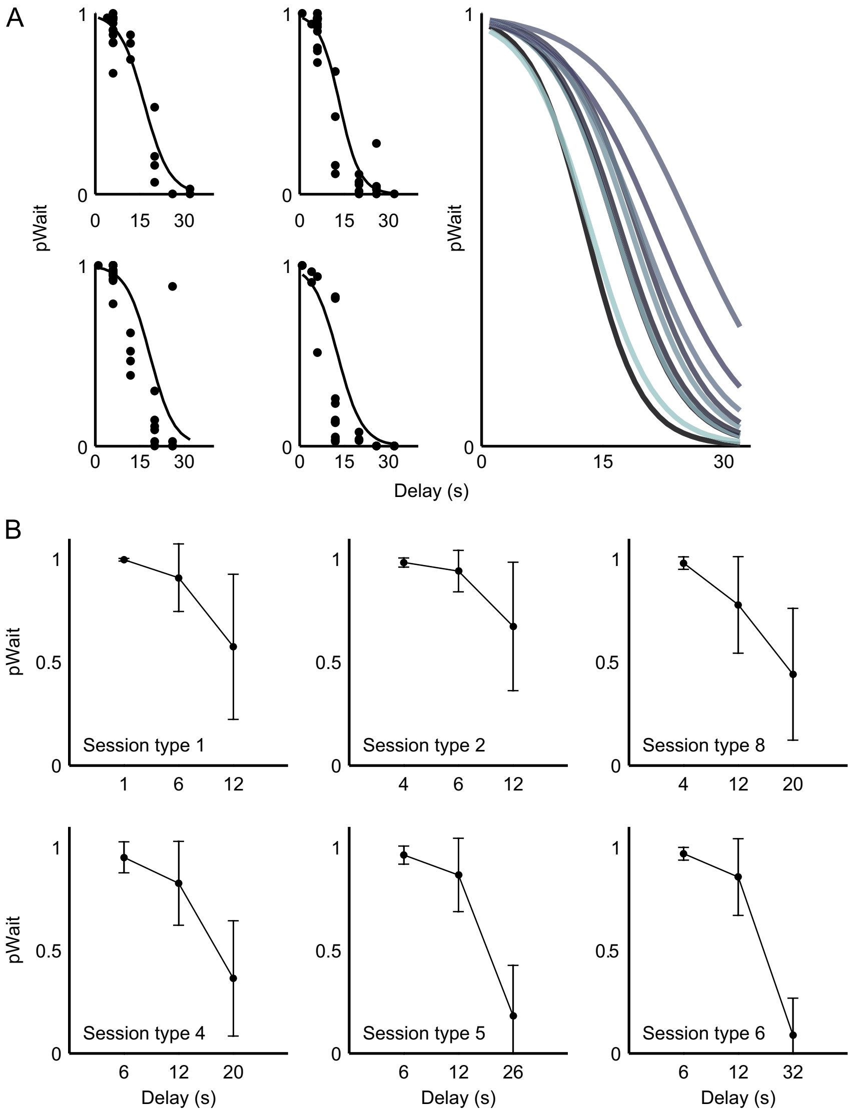

Fig. S1. (*A*) The fraction of trials in which subjects waited for food (p_{wait}) is plotted against delay length across all session types. Sigmoid curves were fit to each subject's data. Data from four individual rats are plotted at *Left*, and the curves for all subjects (*n* =10) are shown at *Right*. Rats varied in their delay tolerances, but in general, p_{wait} decreased with increasing delay. (*B*) The proportion of trials in which subjects waited out the delay period is plotted separately for each session type. Error bars mark the SD (*n* = 10 rats, 4 repetitions per subject of each session type, for 240 sessions of data).

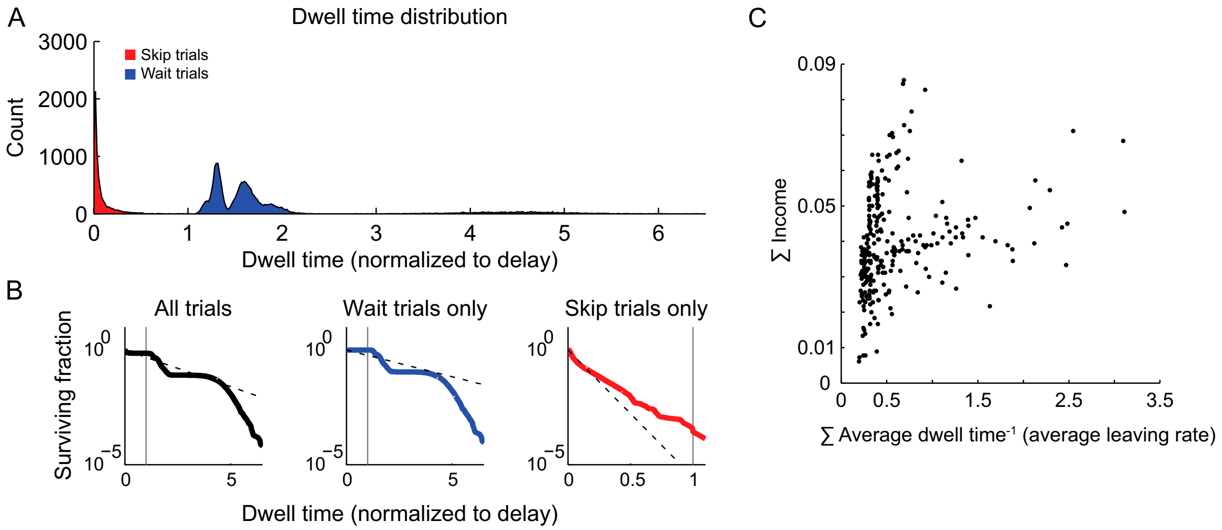

Fig. S2. (*A*) One characteristic of the matching strategy is that dwell times (the time spent waiting at each feeder location) are distributed exponentially. Dwell times for all sessions (normalized to the delay at each location) are plotted, with distributions separated based on whether the subject waited for food delivery (wait trials, blue) or left the site before the delay period had passed (skip trials, red). (*B*) Survivor curves for dwell times (again normalized to the delay at each location) are plotted for all trials, wait trials only, and skip trials only. The vertical gray line in each plot marks the time of reward delivery. The dashed black line in each plot is the empirical survivor curve for an exponential distribution with the same mean as the dwell time distribution. For all three cases, the data-derived survivor curves differ substantially from the curves expected for a matched exponential distribution. (*C*) In matching tasks, the sum of the average leaving rates (the reciprocal of the average visit duration) at all potential food sites typically increases in proportion to the sum of the income across all food sites. We did not observe a clear relationship between leaving rates and total income, suggesting subjects were not guided by a matching strategy on the foraging task.

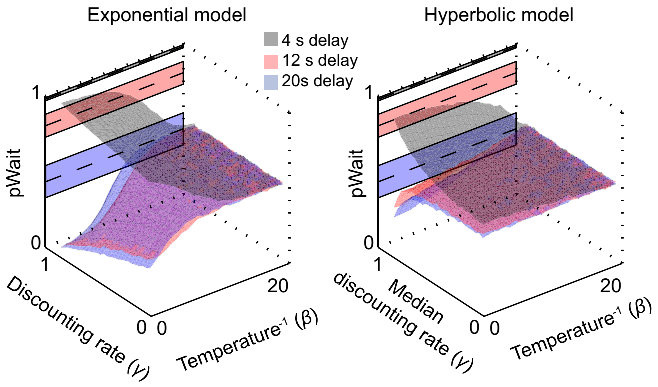

Fig. S3. The behavior of the exponential (*Left*) and hyperbolic (*Right*) Q-learning models is displayed as surfaces in parameter space, with the model-predicted p_{wait} values for each delay plotted in a different color. For comparison, subjects' actual behavior is projected behind the surfaces (mean p_{wait}; shaded regions indicate 95% confidence intervals around the mean).